UNFOLDING ENERGETICS AND PATHWAYS OF PEPTIDES WITH VARYING SECONDARY STRUCTURE MOTIFS

A Thesis
Presented to
The Academic Faculty

by

Hailey Rhea Bureau

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Chemistry and Biochemistry

Georgia Institute of Technology
May 2017

UNFOLDING ENERGETICS AND PATHWAYS OF PEPTIDES WITH VARYING SECONDARY STRUCTURE MOTIFS

Approved by:

Dr. Rigoberto Hernandez, Advisor
School of Chemistry and Biochemistry
Georgia Institute of Technology

Dr. Joseph Perry
School of Chemistry and Biochemistry
Georgia Institute of Technology

Dr. Thomas Orlando
School of Chemistry and Biochemistry
Georgia Institute of Technology

Dr. Stephen Quirk
Global Director of Life Sciences
Kimberly-Clark Corporation

Dr. Jeffrey Skolnick
School of Biological Sciences
Georgia Institute of Technology

Date Approved: March 22, 2017

For my mother.

ACKNOWLEDGEMENTS

I have often heard that it "takes a village" to accomplish anything worth accomplishing. I cannot express how much I now recognize that to be true. First and foremost, I would like to thank my mother for her constant love and support. I am completely convinced that moms really do know everything. Next, I would like to extend my deepest gratitude to my advisor, Prof. Rigoberto Hernandez. Thank you for your patience and guidance throughout the entirety of my graduate career. Of course, I owe a tremedous amount of my success to the Hernandez group, both past and present members. I would like to extend a special thanks to Dr. Caley Allen for constantly offering an alternative perspective to seemingly hopeless situations and for teaching me how to appropriately use "passive-aggressive sprinkles". I would like to express my gratitude for my thesis committee including Prof. Thomas Orlando, Prof. Jeffrey Skolnick, Prof. Joseph Perry, and Dr. Stephen Quirk with special thanks to Dr. Quirk who collaborated with me on several projects within this thesis.

Finally, thank you to all of the friends I have made while at Georgia Tech. I would like to especially thank Johanna Smeekens, Stephen Shiring, Alex Hyla, Erin Gawron, Allison and Mena Aioub, Rachel Stryffeler, Chelsea and Justin Bordley, Brandon Bakr, Alyssa Blake, and Joe Richardson. Your companionship and laughs along the way have helped me stay positive and motivated throughout the years. I am eternally grateful for all of you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

Biomolecules present an almost endless challenge with regard to predicting their properties, structure, and behavior. The present body of work focuses on using computational techniques to characterize and predict the energetics and other observables of peptides unfolding under external forces. This work is composed of two primary aims: to develop efficient computational approaches to address this problem and to apply these methods to specific biochemical problems to learn new chemical biology. In combination, these two aims have led to studies that provide insight into the unfolding pathways and characteristics of several peptide motifs.

In fulfillment of the first aim, we have introduced two new methods to the field of computational unfolding simulations: Adaptive Steered Molecular Dynamics and Full-Relaxation Adaptive Steered Molecular Dynamics. Futhermore, we have benchmarked these algorithms with varying constraints including different peptide secondary structure motifs, solvation models, force field versions, and molecular dynamics integrators.

The second aim of this work is directed toward utilizing those methods to approach complex biochemical problems. We focus on applying our methods to mechanically unfolding several different peptide secondary structure motifs such as helices and β-hairpins. We calculate the energetics of unfolding the peptides, analyze the hydrogen bonding networks, and calculate the interaction energies of several specific residue pairs. Throughout this work, we have collaborated with Dr. Stephen Quirk, an experimental biochemist with the Kimberly-Clark Corporation. As such, both computational and experimental data will be presented and discussed for two of the projects.

Overall, this thesis demonstrates that we have provided one avenue for addressing the challenge of developing computational methods that simulate peptides unfolding under external forces.

CHAPTER I

INTRODUCTION

1.1 Background

Proteins are macromolecules that carry out a variety of functions in cells such as catalyzing reactions, transporting molecules, and cellular signaling. Protein function and behavior has been studied extensively; however, the mechanism of a protein folding into a native state or unfolding from a native state to a random coil is not well understood [1, 2, 3, 4, 5]. Even less is understood about protein misfolding that leads to neurodegenerative diseases [6]. Protein folding is a very complex process and its characterization presents a challenge to both experimentalists and theorists. There have been several experimental methods developed to study protein folding. This includes numerous methods in crystallography, spectroscopy and mechanical probes [7, 8, 9, 10, 11, 12]. Theorists have developed several computational methods, with emphasis on molecular dynamics (MD), to investigate the folding, structure, and properties of proteins [13, 14, 4, 15, 16].

Studying the unfolding pathways of proteins has been shown to also uncover ways to understand the mechanism of folding. Studies of the unfolding of a protein as a means to understand how proteins fold [17]. For example, Anfinsen studied the reversible denaturation of several proteins using urea and β-mercaptoethanol. He determined that the tertiary structure of a protein is determined only by the amino acid sequence. He was the first to suggest that folding is driven by the protein adopting a native structure that has the lowest free energy.

Many of the roles performed by proteins depend on their mechanical properties, including folding and unfolding. The response of proteins to external mechanical manipulation has been studied experimentally by single-molecule pulling experiments using Atomic Force Microscopy (AFM) [7], laser optical tweezers [18], and biomembrane force probes [19]. These types of measurements yield information about the extension of the protein as a function of

the applied force. In general, using these experimental techniques, one obtains the overall view of the protein stretching without any specific, detailed information on the unfolding pathway.

Theoretical approaches have been able to shed some light on the underlying structure and properties of unfolding proteins at the atomistic level. In particular, Steered Molecular Dynamics (SMD), inspired by AFM experiments, has been used to understand the mechanical properties of unfolding or stretching proteins. However, several challenges impede the practicality of using SMD [20, 21, 22, 23, 24]. The immense sampling sizes required to converge the energetics of unfolding present a substantial challenge that must be overcome. Our group has developed an enhanced-sampling method, Adaptive Steered Molecular Dynamics (ASMD) to circumvent this problem. Additional improvements upon the original ASMD method have resulted in the development of Full-Relaxation ASMD (FR-ASMD), Partial-Relaxation ASMD (PR-ASMD), and Multiple-Branched ASMD (MB-ASMD).

1.2 Motivation

Understanding the mechanistic pathways involved in protein folding has long been a subject of interest for both theorists and experimentalists. Knowledge of the unfolding or stretching of peptides by mechanical perturbations provides insight into their energetics, dynamics, and stabilities in various environments. Unfolded proteins that accumulate in the endoplasmic reticulum of cells are known to induce a signal-transduction response, referred to as Unfolded Protein Response (UPR) [25]. UPR has been linked to a number of aliments such as various cancers, heart disease, neurodegenerative diseases, and immune disorders. Determining the energetics of the unfolding pathways of numerous types of peptide motifs allow for the potential development of methods to control misfolding events that are known to lead to such degenerative diseases [26]. Additionally, understanding the energetics of protein unfolding has potential, but far-reaching implications for the field of drug design and delivery.

1.3 Thesis Objectives

The aims of this thesis are to develop computational methods to efficiently simulate the process of a peptide unfolding under external forces and to obtain the relative energetics

of differing secondary structure motifs to uncover the driving mechanisms contributing to the overall stability of the peptides. In order to investigate the mechanisms of protein stability, several different peptides have been studied using different variations of SMD developed in our group. By applying these methods we have gained a better understanding of the dynamics of small single-motif peptides. This has been accomplished by investigating the unfolding energetics and accompanying structural changes of α-helices, β-hairpins and multi-motif peptides.

1.4 Thesis Structure

In Chapter 2, the computational methodology used throughout Chapters 3 through 5 are presented. This includes a discussion of several non-equilibrium sampling methods with a focus on SMD. The section also presents methods developed by the Hernandez group including ASMD, FR-ASMD, PR-ASMD, and MB-ASMD. Additionally, the methods used to calculate the weighted observables, such as hydrogen bonding patterns and residue pair interaction energies are also presented.

Chapter 3 focuses on the energetic trends and structures obtained during the unfolding pathways of helical peptide motifs. It is a compilation of published work [27] on the unfolding of ALA_{10} under varying solvent conditions co-authored with Dale Merz Jr., Dr. Eliezer Hershkovits, Dr. Stephen Quirk and Prof. Rigoberto Hernandez and work currently in preparation on the unfolding behavior of helical peptides in three versions of the CHARMM force field [28], using different families of potentials (i.e. AMBER and CHARMM) [29], and the linear trends that emerge during the unfolding of alanine-rich peptides of varying lengths [30].

Chapter 4 contains three different studies of the dynamics of several β-hairpins. The studies in this chapter are adapted from published work and two manuscripts currently in preparation. First, published work [31] comparing the unfolding energetics, hydrogen bonding patterns and residue pair interaction energies between chignolin and trpzip1 co-authored with Dr. Stephen Quirk and Prof. Rigoberto Hernandez will be presented. Next, work in preparation focusing on a joint experimental and theoretical mutation study of

trpzip1 [32] is discussed. Lastly, another study in preparation focusing on the differences in the energetics of unfolding β-hairpins depending on the family of potential used for the molecular dynamics [29].

Investigations of the unfolding dynamics of multi-motif peptides are presented in Chapter 5. This includes two joint experimental and theoretical studies currently in preparation. The first study focuses on the energetics and formation higher-order structures of three mutations of a polyproline helix [33]. The second study centers around a mutation study aimed at modulating the stability and function of an antimicrobial peptide [34].

Chapter 6 elaborates the overall conclusions of this thesis and the significant findings of Chapters 3 through 5.

CHAPTER II

METHODOLOGY

One of the advantages of using MD methods is the ability to simulate systems with an incredible number of possible conformations. Biomolecular systems inherently contain many such possible conformations due to their flexible backbones and side chains. In order to accurately and efficiently capture those conformations, numerous MD sampling methods have been devised.

2.1 Non-equilibrium sampling methods

Free energy perturbation (FEP) is a statistical mechanics method developed by Zwanzig for the simulation of nonpolar gases [35]. It is a method for calculating differences in free energy from MD and Monte Carlo simulations. The change in free energy associated with a transition from an initial to a final state is characterized by the Zwanzig equation [35]:

$$\Delta F(1 \rightarrow 2) = F_2 - F_1 = -k_B T \ln \Big\langle \exp\Big(- \frac{E_2 - E_1}{k_B T} \Big) \Big\rangle_A \tag{1}$$

where T is the temperature, k_B is Boltzmann's constant, and the brackets indicate that the average is performed from a simulation of the system in state 1. As a simulation of the initial state is performed, every instance a different configuration is accepted, the energy for state 2 is calculated again. This algorithm may be applied to a number of different types of systems. For example, states 1 and 2 could refer to different geometries of the system. In any case, a profile of the free energy changes along one or more reaction coordinates is obtained for the transition. This constructed free energy landscape is referred to as a PMF. A limitation of FEP calculations is that the PMFs will converge only when the change between the two states is relatively small. As such, the algorithm requires division of the reaction coordinate into segments which are referred to a "windows" and can be computed independently; hence, increasing the speed at which computations can be achieved. The

FEP methods has been used in combination with MD and quantum mechanical approaches to compute a several biological processes including ligand binding, solvent effects of reactions and enzymatic reactions.

Building on the achievements of the FEP methodology, additional methods were developed in order to improve the sampling of systems with large energetic barriers, as can be the case with chemical reactions or the transitions of a biomolecule between unstructured and native states. Umbrella sampling, developed by Torrie and Valleau, was developed to overcome such sampling limitations [36]. Modifications to the potential function, such as a biasing force, allow the system of interest to surmount the large barrier. This method is called "umbrella" sampling because the reaction coordinate is divided into windows which overlap. The overlap of the windows can be modified in order to fit the constraints of any given system. The bias applied to the potential energetically connects the separated regions able to be sampled. Once a reaction coordinate has been determined, segmented, and the bias has been applied to the potential, it is then necessary to sample within the window in order to obtain a converged probability distribution function. The probability distribution function is then unbiased and can be inverted to calculate the free energy profile.

One such method for obtaining the free energy profile is Weighted Histogram Analysis Method (WHAM) [37]. WHAM was developed as an extension of the Multiple Histogram Technique [38, 39]. It is a method for extracting the free energy profile from sampling schemes that involve using multiple reweighing procedures. The guiding principal of WHAM relies on having a discrete number of states. A histogram constructed from discrete bins can be calculated, which yields the relative probability of observing the events. From these probabilities, the weighted free energy profile along with other desired observables can be calculated.

2.2 Steered Molecular Dynamics (SMD) and the Jarzynski Equality (JE)

One successful theoretical method used to predict the free energy profile for a number of different types of systems is SMD [40, 41, 23, 24]. This method was used to find the free energy of: the unfolding a peptide [23], transitions between different states of a molecule

[23, 24] and forcing a molecule through a channel [40], and to understand the dissociation of a salt compound in a solvent [42]. SMD is designed around the following procedure: one end of the peptide is restrained using a harmonic potential of the form:

$$h_\lambda(r) = \frac{k}{2}(\xi(r) - \lambda)^2, \tag{2}$$

while the other end is pulled with constant force or velocity until the peptide is completely unfolded. SMD is often used in conjunction with the JE [43, 44, 45, 46]:

$$G(\xi_t) = G(\xi_0) - \frac{1}{\beta} \ln \left\langle e^{-\beta W_{\xi_t \leftarrow \xi_0}} \right\rangle_0 , \tag{3}$$

to average the work distribution and obtain the PMF. The JE is a non-equilibrium work theorem that relates equilibrium observables to non-equilibrium trajectories. SMD was employed in the calculation of the PMF of the unfolding of decaalanine, a small α-helical peptide [24]. This method was found to be comparable to umbrella sampling using WHAM for analysis. However, SMD in conjunction with the JE provides a much simpler method for calculating the PMF in comparison with WHAM [23].

Although SMD is a useful method, it can be very computationally expensive because a large number of trajectories are required to converge the PMF for long distance pulling paths. One of the main drawbacks of SMD is that the logarithm of an average of exponentials, referred to as the "exponential average" of the work, is much lower than the average of the work. Therefore, the trajectories with the greatest contribution in the Jarzynski average are rare events. Finding these rare events requires averaging over a very large sample space.

There have been several attempts to reduce the number of trajectories required for the correct estimation of the free energy. The use of backward as well as forward trajectories have lead to improved convergence in the averaging of the work [47, 48], which is further enhanced through the use of the maximum likelihood estimator (MLE) [49, 50]. Multistage methods, such as Kofke's multistage approach [51], multistep trajectory combination (MSTC) [52, 53], our own ASMD method [54, 55, 56], along with and several others [57, 58, 59] have also been developed in order to decrease divergence of the work functions of trajectories. These methods share a similar strategy in that they are all based

on segmenting the pulling path into several smaller stages and independently calculating the Jarzynski average along these stages. The advantage of this segmentation lies in the fact that a smaller stage leads to a smaller span of trajectories which can be more readily fully sampled despite the possibility of rare low-work trajectories.

Naïve ASMD and FR-ASMD follow the second of these strategies. In both of these methods, the major advantage involves the contraction of trajectories at the end of a stage to structures closer to those representative of the equilibrium distribution of the peptide constrained to the particular extension. In naïve ASMD, a single structure is chosen from those in the numerically determined non-equilibrium ensemble according to which had the associated work closest to the Jarzynski average [54, 55]. This is further explained in Sec. 2.3. In FR-ASMD, all of the trajectories are relaxed for some additional integration under the constrained extension [52, 55]. As this additional constrained non-equilibrium path requires no work on the system, the Jarzynski average is unaffected through the relaxation stages of FR-ASMD. A complete discussion of this "enhanced" ASMD along with other methods can be found in Sec. 2.4.

2.3 Adaptive Steered Molecular Dynamics (ASMD)

The main method for obtaining the PMFs within this work is the ASMD method [54]. The efficiency and accuracy of the algorithm have been benchmarked for vacuum [55], implicit solvent [27] and explicit solvent [56] conditions for a small peptide (e.g., ALA_{10}), and it has effectively been used to reveal the energetics of a large peptide, Neuropeptide Y (NPY) [54]. For the work in Chapter 4, ASMD was used, for the first time, with β-hairpin motifs. In light of new challenges with NPY, the complete unfolding pathways is obtained in Chapter 5 along with a mutation study of Enterocin 7B. In the ASMD method, the overall reaction coordinate is divided into segments and the PMF is calculated over each segment within an SMD-like stage using the JE [43, 44, 45, 46]. The reaction coordinate [60] is arbitrarily defined as the z-axis, along which the peptide is oriented from end-to-end and stretched in one direction.

The n_s stages are partitioned across the steering coordinate which, in the present case,

is the end-to-end distance r_{ee} of the protein. We denote the set of $n_{\mathrm{s}}+1$ endpoints as $\{r_{\mathrm{ee},j}\}$ where $j=0$ corresponds to the initial position. Thus far, these end points have been taken to be equally spaced across the r_{ee} domain, but their positions could be placed at varying separations along the domain without affecting the formal accuracy of the method. Similarly the stretching coordinate is pulled at a constant velocity, v_{s}, such that the auxiliary particle position is $r_{\mathrm{ee}}(t) = r_{\mathrm{ee},0} + v_{\mathrm{s}}t$. For a given position $r_{\mathrm{ee}}(t)$ in the set $(r_{\mathrm{ee},j}, r_{\mathrm{ee},j+1})$, the average work can be obtained through the Jarzynski equality as

$$\overline{W}(r_{\mathrm{ee}}(t)) = \overline{W}(r_{\mathrm{ee},j}) - \beta^{-1} \ln \left\{ \sum_{i=1}^{N} e^{-\beta W_i\left(\xi_t^{(i)}\right)} \right\} , \tag{4}$$

where $\xi^{(i)}$ is the i^{th} trajectory in the non-equilibrium ensemble stretched from $r_{\mathrm{ee},j}$.

In between stages, the ensemble space covered by the final structures of the previous stage is contracted according to criteria —whose specifics are determined by the type of ASMD being performed— that add no work to the system. In naïve ASMD, the contraction is imposed by determining the trajectory for which the work is closest to the PMF value. The initial stretch of this structure, at the beginning of each stage, using a distribution of (stochastic) thermal forces leads to an ensemble of states near a local equilibrium in the vicinity of the initial starting configuration is obtained before a significant deviation of the pull. (For example, at v_{s} equal to 1 Å/ns and 1fs time steps, a 0.01 Ådeviation requires 10,000 steps.) Such a local equilibrium ensemble is obtained at the beginning of a standard SMD simulation where it is considered to be safe to assume that it is a global equilibrium ensemble. Thus the main assumption in its use at subsequent stages is that the local equilibrium ensemble is sufficient to capture the global equilibrium ensemble for a fixed displacement [55, 56]. When this is not satisfied, we would then need to extend the initial configurations to access the global equilibrium ensemble such as is possible through FR-ASMD [27, 55] or MB-ASMD [61]. (As this work uses only naïve ASMD and not any of the other recent variants employing alternate contraction criteria, such as FR-ASMD [27, 55] or MB-ASMD [61], we refer to the method simply as ASMD throughout.) The coordinates of the final point of this trajectory are used as the starting point for the next stage. In that next stage, the velocities for each trajectory are distributed based on random

number sequences.

After the PMF is calculated along the first segment, the final value in the work that most closely matches the final value in the PMF in that stage determines which coordinates and velocities are selected for all trajectories in the subsequent stage. A new ensemble of distinct trajectories is generated in the latter because each trajectory is pulled in the presence of a Langevin bath with different random number sequences. The use of the so-called JE-criterion to contract the sampling space between stages results in faster convergence of the PMF and reduces the overall number of trajectories that must be calculated in comparison with SMD results [54].

The work performed on a peptide during these pulling simulations, when performed infinitely slowly, is equal to the Helmholtz free energy difference. With Jarzynski averaging, the velocity of pulling can be increased two to five orders of magnitude (to sample speeds of 100, 10, and 1 Å/ns, for example), while still permitting one to obtain equilibrium free energy differences from finite-time measurements. The benefit of using ASMD is consequently increased sampling efficiency in the sense that fewer trajectories and reduced computational resources are employed to obtain a more highly converged PMF than comparable SMD sampling.

An example of a PMF obtained using ASMD can be seen in Fig. 1. This is the PMF of the forced unfolding of chignolin, a 10-residue β-hairpin further presented in Chapter 4. As the peptide is pulled within a stage, the work values spread farther apart. At the end of each stage, all of the trajectories are contracted to a single point. The theoretical basis of the ASMD method, including discussion of the convergence criteria, may be found in our previous work [54, 55]. Operationally, we ensure that within a given stage the work trajectories spread no more than 3 $k_{\mathrm{B}}T$ while retaining constant stage lengths for simplicity. We also vary sampling parameters to obtain overall convergence in the final pulling energy to within 1 $k_{\mathrm{B}}T$ (≈ 0.6 kcal/mol).

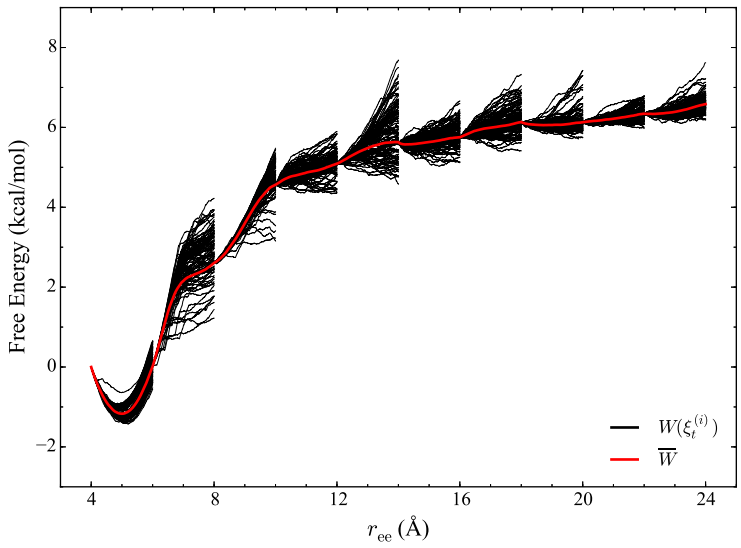

Figure 1: The work energy trajectories $W\left(\xi_t^{(i)}\right)$ (in black) and the PMF $\overline{W}$ (in red) obtained using ASMD through the adaptive stages of Eq. 4. Although a cartoon could have been used, the illustration shown here corresponds to the data obtained for chignolin with 100 tps at 0.1 Å/ns which is further discussed in Chapter 4.

2.4 Enhanced ASMD

Though naïve ASMD provides many advantages over SMD with regard to convergence efficiency, there are several different ways to divide the reaction coordinate and different ways to perform the contraction between each stage. These differences in the division of the reaction coordinate and contractions leads to further efficiencies in the convergence of the PMFs.

2.4.1 Full-Relaxation ASMD (FR-ASMD)

One of the improvements to the ASMD method for obtaining faster convergence of the PMF is through the use of intermediate stages for minimization and equilibration. FR-ASMD [27, 55] invokes different choices for the segmentation of the reaction coordinate and the contraction in between each stage. It is similar to naïve ASMD in that it requires one to break up the reaction coordinate into segments. During the corresponding stages, the peptide is pulled from some initial distance to a final distance. However, between pulling stages, a constrained equilibration stage is used to construct the initial configuration for the

next pulling stage. During the constrained-relaxation stage, the end points of the peptide are fixed to their initial positions. As no work is done on the peptide during the constrained relaxation, the Jarzynski average is unaffected by this relaxation. In relaxing the structures with large energy deviations from the minimum energy structures, however, strains that could have skewed the subsequent distributions are removed. This contraction of the initial sample space for a given pulling stage has a similar effect as the more extreme contraction (to a single structure) implemented in naïve ASMD, but without incurring any error in the free energy average. However, it comes at the price that one must integrate the system for a time longer than the relaxation times of the slowest mode orthogonal to the fixed pulling coordinate.

2.4.2 Partial-Relaxation ASMD (PR-ASMD)

Similar to FR-ASMD, PR-ASMD is another improvement to the ASMD method for obtaining faster convergence of the PMF. PR-ASMD utilizes intermediate stages for redesigning the solvation box dimensions, hence reducing the number of atoms in the simulations. The reaction coordinate is divided into an predefined number of stages as in an ASMD experiment. Between each "pulling" stage the final coordinates of the peptide that would have been selected for continuation in the next stage in a typical ASMD simulation are striped of solvent and resolvated in a smaller water box. Each simulation stage is performed in a solvent box that redesigned to accommodate the entire length of the peptide surrounded by a generous margin of solvent. A series of equilibrations ensure that the dynamics of the peptide are not egregiously affected by the resolvation process. After the peptide is resolvated, the system is equilibrated under an NPT ensemble with the peptide structure restrained. Next, the system is again equilibrated using the NPT ensemble; however, only the Cα of the termini are restrained. Lastly, an equilibration is performed using NVT conditions again with the with the Cα of the termini restrained. In order to ensure a distribution of starting structures, a predefined number of initial configurations of the system is randomly selected from the last few nanoseconds of the equilibration. Those starting structures are used in the next pulling stage. This process is repeated until the end of the reaction coordinate is

reached. As no external pulling force is applied to the system during this time, no work is being performed on the system during the equilibration stages. Hence, the overall unfolding pathway of the peptide is not being unduly manipulated by the resolvation and subsequent equilibration of the system.

2.4.3 Multiple-Branched ASMD (MB-ASMD)

One criticism of naïve ASMD is the possible lack of sampling along multiple attractive basins, that may exist, along the unfolding coordinate of a peptides. This feature may be exaggerated in the case of larger proteins which have unfolding free-energy landscapes that can access multiple pathways. In order to address this potential shortcoming, MB-ASMD was developed jointly by the Keyes and Hernandez groups [61]. In this approach, the reaction coordinate is, again, divided into segments. Multiple basins along the free energy profile are "selected" as potential unfolding pathways instead of selecting only one dominant pathway as is standard in ASMD. In ASMD the selection criteria is based on which trajectory has a final work value closest to the value of the JA at the end of a stage. In MB-ASMD, each trajectory is assigned a probability based upon one of the following criteria:

1. the probability of the trajectory having absolute work values near the JA

2. the exponential average of the negative work

No limit is imposed on how frequently a trajectory is selected as the initial coordinates for the subsequent stage and is based only on the distribution of the work trajectories. However, the more times a trajectory is selected, the fewer different configurations . It was observed that $\approx 50\%$ of the structures obtained at the end of a stage are not selected as initial structures for the next stage. There are several coordinates that are selected only one time and a few are selected multiple times.

This work was benchmarked using decaalanine under vacuum conditions in order to be comparable to the benchmarking of SMD [24] and ASMD [55]. It was found that MB-ASMD does provide efficient construction of the PMF in comparison to SMD and naïve

ASMD. In particular, using the exponential average of the negative work selection scheme yields convergence to the exact (i.e. reversible) PMF at using a pulling velocity that is 100 times faster than the known reversible pulling speed limit. Though it should be noted that decaalanine is a simple model system, likely with few potential unfolding pathways. Nevertheless, this work provides a useful starting point for a comparison. In future work, MB-ASMD will be tested larger proteins that have more numerous potential unfolding pathways.

2.5 Calculation of weighted observables

In addition to the PMFs, other observable quantities can be calculated and compared in order to further probe the underlying causes of stability of a peptide. For example, these observables can be hydrogen bonding patterns or interaction energies of specific pairs of residues. Expectation values of observables along the unfolding (stretching) pathway are obtained using the weights for the work associated with the non-equilibrium paths [55, 60, 62]. In order to calculate the hydrogen bond patterns during the unfolding pathways, let the sets S_1 and S_2 contain the indices for the positions of the selected atoms and/or collective variables in $\vec{\xi}$. The average number of hydrogen bonds between the two sets can then be written as

$$\langle N_{\mathrm{H}}(S_1, S_2)\rangle_t = \frac{\sum_{i=1}^{N} \hat{N}_{\mathrm{H}}(S_1, S_2) e^{-\beta W_i\left(\xi_t^{(i)}\right)}}{\sum_{i=1}^{N} e^{-\beta W_i\left(\xi_t^{(i)}\right)}} \,, \tag{5}$$

where the instantaneous number of hydrogen bonds between the sets is

$$\hat{N}_{\mathrm{H}}(S_1, S_2) = \sum_{k\in S_1, l\in S_2}{}' \hat{N}_{\mathrm{h}}(\xi^{(k)}, \xi^{(l)}) \,, \tag{6}$$

where $\hat{N}_{\mathrm{h}}(\xi^{(k)}, \xi^{(l)})$ is 1 if $\xi^{(k)}$ and $\xi^{(l)}$ are hydrogen bonded positions and 0 otherwise, and the prime in the sum excludes the case that $k = l$. The two positions are identified as hydrogen bonded when the donor and acceptor atoms are within 3.0 Å of each other and when the angle formed between them is less than or equal to 40°. Numerically, the number of hydrogen bonds is obtained using the VMD Plugin, HBonds [63], and summed through Eqs. 5 and 6 using an in-house python script. Let S_{P} and S_{W} denote the sets of

oxygen atom positions in the peptide and water solvent, respectively. The average number of intrapeptide hydrogen bonds is determined by $\langle N_{\mathrm{H}}(S_{\mathrm{P}}, S_{\mathrm{P}})\rangle_t$, and the average number of hydrogen bonds between peptide and solvent is determined by $\langle N_{\mathrm{H}}(S_{\mathrm{P}}, S_{\mathrm{W}})\rangle_t$.

Similarly, the average interaction energy between residues indexed in sets S_1 and S_2, can be obtained from

$$\langle E_{\mathrm{int}}(S_1, S_2)\rangle_t = \frac{\sum_{i=1}^{N} \hat{E}_{\mathrm{int}}(S_1, S_2) e^{-\beta W\left(\xi_t^{(i)}\right)}}{\sum_{i=1}^{N} e^{-\beta W\left(\xi_t^{(i)}\right)}} , \tag{7}$$

where

$$\hat{E}_{\mathrm{int}}(S_1, S_2) = \sum_{k\in S_1, l\in S_2}{}' \hat{E}_{k,l}^{\mathrm{pw}} , \tag{8}$$

$\hat{E}_{k,l}^{\mathrm{pw}}$ is the interaction energy between the particular residues labeled by k and l, and the prime in the sum excludes the case that $k = l$. The latter is obtained from the sum of the pairwise additive energies in the chosen force field. For example, $\langle E_{\mathrm{int}}(S_{\mathrm{Trp4}}, S_{\mathrm{Trp9}})\rangle_t = \langle E_{\mathrm{int}}(\mathrm{Trp4}, \mathrm{Trp9})\rangle_t$ is the average interaction energy between the single residue Trp4 and the single residue Trp9. The VMD plugin NAMDEnergy [63] has been used to calculate interaction energies between residue pairs for each peptide and summed through Eqs. 7 and 8 using an in-house python script.

CHAPTER III

CHARACTERISTICS OF THE UNFOLDING PATHWAYS OF HELICAL PEPTIDES

3.1 Introduction

This chapter details four different studies preformed using helical peptides as the motif of interest. The chapter is outlined as follows: The four helical model systems that have been investigated are described in Sec. 3.2. A detailed description of the unfolding pathways in ALA_{10} in various solvent conditions is presented in Sec. 3.3.1. Benchmark data on the performance of ASMD, with two helical motifs, using different MD potentials is discussed in Secs. 3.3.2 and 3.3.3. Next, simulation results of the elongation of several alanine-rich α-helices will be discussed in Sec. 3.3.4. Finally, in Sec. 3.4, we conclude by summarizing the overall trends and dynamics observed for helices undergoing induced mechanical unfolding.

3.1.1 Implicit versus explicit solvent models

The numerical determination of the free energy for unfolding large peptides remains a major computational challenge. The requisite integration time is long because the ends of the protein must be pulled to a very long distance at relatively slow pulling speeds. The computational requirements are further exacerbated by the inclusion of an explicit solvent box that has to be large enough to contain the unfolded and folded protein structures with room to spare for solvent shells. As the end-to-end length of the protein grows, the number of required water molecules also grows, leading to ever larger computer memory requirements. These costs can be mitigated through the use of implicit solvent models, but there has been much debate over how well they can mimic explicit solvent effects [64].

Skinner and coworkers, for example, found that the environment plays a critical role in the structure formation of the putative helical segment in rat and human amylin, but the question remains as to whether the solvent must be specified at all-atom resolution in order to obtain such effects in every case [65, 66]. The early work of Paci and Karplus [67],

assumed that an implicit solvent suffices in studying the structure and energetics for pulling titin. However, subsequent work on small alkanes [68] and ALA_{10} [69] suggests that the water model representing the solvent needs to be considered at least at a coarse-grained level so as to obtain the correct energetics and structure along the pathway. In our recent work [56], we found that the energetics and the hydrogen-bonding pathway along the stretching of ALA_{10} —whose structures are illustrated in Fig. 2,— was quite different between vacuum and explicit solvents. However, Gumbart and coworkers [70] recently found a PMF for ALA_{10} in explicit water that differed from our previous result. As such, the determination of the correct pathway presents a significant test of the ASMD approach as well as the use of an implicit solvent model. Surprisingly, in the present work, we found that the overall pathway, in which internal hydrogen-bonds are broken and supplanted by contact to the solvent, is mostly recovered by the implicit solvent.

The use of implicit solvent models has gained popularity because they provide the possibility for reducing a large number of degrees of freedom (as it effectively coarse-grains the solvent away) and thereby reduce the computational cost significantly. They have been seen to yield relatively accurate predictions of protein native structure [71, 72] and even folding pathways of short peptides [73, 74, 75]. Reference [76] provides a recent comparison of the relative accuracy in the energetics of several implicit solvent models. Earlier numerical experiments of a β-hairpin folding by Zhou and coworkers [77] found that the folding structure depended heavily on the choice of the GBSA force field; only the Amber96/GBSA force field gave the known native structure. They also found that the implicit solvent models tend to elongate helices and reduce the stability of the β-hairpins relative to the native structure. Nilsson and coworkers [78] performed MD studies at four different temperatures in ten different implicit solvent models implemented in CHARMM. They found that a variant of the GBIS model —namely Generalized Born using molecular volume (GBMV II)— was the best choice for capturing the solvation energetics in unfolding and folding simulations. There now exists ample evidence [71, 79] that implicit models can be used to obtain accurate near-equilibrium structures as long as a surface-dependent term is included in the representation. The question explored in the present work is whether this is also sufficient

to obtain the correct PMF and structural pathway through SMD-like calculations.

3.1.2 Benchmarking ASMD using different CHARMM potentials

The energetic interactions within biomolecules and their role in structure-function relationships can be partially understood through the determination of their associated work functions —thermodynamics— along the characteristic path between the interacting components. Computational methods for obtaining these work functions include, but are not limited to, replica exchange molecular dynamics [80], FEP [35], and SMD [23, 24, 22]. We have focused on SMD approaches in combination with the JE [44] because it offers the possibility of general applicability as long as the sampling of non-equilibrium trajectories converges.

The SMD method has been seen to accurately describe the free energy profile of a biological system transitioning between stable states along a predefined steering pathway (*e.g.*, a predefined peptide unfolding reaction coordinate) [23, 24, 70, 22]. Unfortunately, as we found in our previous NPY study [54], the convergence of the non-equilibrium trajectories can be slow and worsens as the overall extension along the non-equilibrium reaction coordinate is increased. As the extension grows, many of the underlying non-equilibrium trajectories wander increasingly far from the dominant unfolding pathways, requiring significantly higher work, and hence contribute little to the average of the exponential work in the Jarzynski Equality. This leads to the need for an extraordinarily large and increasing number of non-equilibrium work trajectories that must be calculated in order to converge the resulting potential of mean force (PMF). Our method, ASMD, rectifies these obstacles by dividing the reaction coordinate into a series of smaller segments over which the work distribution is narrower and hence requires many fewer non-equilibrium work trajectories to converge. We have previously demonstrated the accuracy of this approach for the mechanical unhinging of NPY [54], for the mechanical unfolding of the hydrophobic-homopolymer ALA_{10} [27, 55, 56], and the unfolding of two small β-hairpin peptides [31].

The quality and reliability of these simulations is largely dependent upon the accuracy of the underlying empirical force field (FF) of choice. There are many available choices for the

all-atom, additive potential functions, such as the CHARMM [81, 82, 83], Amber [84, 85], and OPLS-AA FFs [86, 87]. These empirical FFs are continuously being improved, resulting in more accurate quantitative energy calculations ranging from structure information to thermodynamic properties, while unfortunately creating questions as to which version is most accurate for a particular observable. In particular, recent refinements to additive FFs have focused on the reparameterization of the torsion angles and corresponding degrees of freedom.

For example, ff99 was shown to over stabilize the α-helical secondary structure conformation [88]. Through a reparameterization of the backbone ϕ and ψ torsions obtained by fitting the energies of glycine and alanine tetrapeptides from high level ab initio quantum mechanical calculations, Simmerling and coworkers were able to make improvements and developed ff99SB [89]. More recently, ff14SB included additional corrections to the backbone and side-chain torsion potentials [90]. The reparameterization of the original OPLS-AA FF involved generating a set of energies for all amino acids on the basis of geometry optimizations and single point LMP2/cc-pVTZ(-f) calculations [87]. Afterward the backbone ϕ and ψ angle parameters were refit to this data. The CHARMM FF potential reparameterized the internal parameters associated with the peptide backbone via the CMAP and dihedral terms [91]. Through the above mentioned series of very subtle modifications to the FFs potential energy functions and corresponding parameters, significant changes in the computed biological behaviour with respect to the overall stabilization or destabilization of the peptide system can occur. An additional challenge to the accuracy of FF based methods is that forty-three percent of the proteins in the Protein Data Bank (PDB) are of unknown function and therefore have little or no experimental data to compel and validate changes to FFs [92]. Thus, the uncertainty in the choice of FF makes it difficult to perform a single calculation of the potential of mean force without consideration of the other FFs.

The aim of this article is to clarify which version of the recent CHARMM family of FFs is most appropriate for the use of ASMD, and to provide additional evidence for the efficacy of the approach through the determination of the potential of mean force for a peptide not previously addressed using ASMD.

Specifically, we focus on a *de novo* designed amphiphilic 18-residue peptide, 1PEF (EQLLKALEFLLKELLEKL) that possesses a high degree of helicity and self-associates into hexamers in aqueous solvent [93]. It forms aggregates and is a possible model for mimicking the peptide aggregation processes seen in many neurodegenerative diseases, such as Alzheimer's and Parkinson's [26]. However, the time scales needed to characterize peptide aggregation and self-assembly can be long, and thereby pose a challenge for atomic-scale molecular dynamic simulations [94]. Nevertheless, the structure and stability of the single 1PEF peptide is germane to the larger complex. We address it using ASMD so as to provide a physically realizable alternative to ALA_{10} that is structurally larger and contains significant helical character.

In this work, we performed a rigorous evaluation of the ASMD methodology with several established CHARMM FFs: CHARMM22 (c22) [81], CHARMM27 (c27) [82, 95] and CHARMM36 (c36) [83, 91, 96]. Specifically, we compared and analyzed the energetics and conformational sampling of two small α-helical peptides, ALA_{10} and 1PEF, along a predefined reaction coordinate to improve the general understanding of the FF parameters and benchmark the effects of our method on relevant conformational changes and the corresponding mechanical unfolding energetics. In particular, we confirmed that our ASMD methodology correctly captures, through efficient sampling and averaging, the previously reported π-helical structural artifacts observed for c22, the α-helical bias reported for c27, and the corrections reparameterized in the CMAP term for c36.

In many currently available all-atom additive FFs, the all inclusive many-body effects is generally missing, and these effects make a large contribution to the total energy in the condensed phase [97]. For the helix-to-coil transition, this enhancement of the total energy may be manifested from stronger hydrogen bond formations [96]. as a consequence of charge transfer effects [98] or electronic polarization [99]. Through the use of a single family of MD potentials, we will show how ASMD is sensitive to the underlying FF, but improves systematically with the evolution of the FF parameterization; both with respect to the mechanical-unfolding energetics and the pathway of each α-helical peptide system.

The first CHARMM FF of interest to this study is c22 which differed significantly from

it's predecessor EF2 [100]. The c22 FF was truly the first modern CHARMM FF, and was parameterized to include an explicit TIP3P water model [101], nucleic acids [102], proteins [81], and lipids. The c22 FF has been extensively used to study the mechanical unfolding of peptides in explicitly represented water molecules. For example, Schulten and coworkers in 1999 examined the force-induced unfolding of the α-helical domain IGB and the β-sandwich domain CAD2 [21]. Likewise, our original NPY [54] and ALA_{10}[55, 56] ASMD simulations were carried out with c22. However, despite its broad adoption, the c22 FF has been seen to have some shortcoming, including its improper handling of protein conformations leading to the incorrect prediction of π-helical structures [103].

To address these issues in c22, improvements in the backbone energetics were undertaken in the form of a two-dimensional dihedral energy correction map (CMAP) that could be applied to the $\phi\psi$ functions of c22 [82]. The CMAP grid correction was applied to all residues in the c22 FF, except glycine and proline for which different CMAP corrections were calculated. c27 has been used to study the dynamics of various biological systems, ranging from structural-function studies of proteins [104, 105, 106], long scale MD simulations [107], to SMD simulations [108].

Even though the CMAP correction to c22 resulted in better agreement with NMR measurements, c27 has been shown by several groups to have a bias toward α-helical conformations [68, 109]. This bias is largely due to relatively small inaccuracies in the backbone potential [68, 110, 111, 112]. To correct this issue, CMAP was reparameterized in the c36 FF by Mackerell and coworkers [113]. The grid correction map was adjusted to provide a better fit to experimental NMR structure data for the peptides ALA_5 and Ac-$(AAQAA)_3$-NH_2 within common residues, and to the QM energy surfaces computed with a higher level of theory for the glycine and proline residues. Additionally, the side-chain dihedral parameters χ_1 and χ_2 for each amino acid were fitted to QM surfaces [91, 114]. Recently, Gumbart et *al.* characterized the thermodynamics of decaalanine folding in water using both the c27 and c36 FFs. They determined that c36 agreed well with previously reported PMFs and significantly improved agreement with helix-formation experiments [70].

3.1.3 Comparing the performance of CHARMM36 and Amber16 using the helical motif

One of the major challenges in the field of computational biomolecular simulations is the development and benchmarking of different potential energy functions. Throughout the years, several different families of potentials have been developed as introduced in Sec. 3.1.2. The two potentials of interest here are the CHARMM and Amber potentials. As such, numerous studies have been conducted on the differences between the performance of various aspects of the potential functions. Studies by Scheraga and coworkers focused on comparing the performance of the potentials using small peptides as the model system. Their first study focused on the comparison of early versions of the CHARMM, Amber and ECEPP potential energy functions using conformational predictions for the peptide (ASN-ALA-ASN-PRO)$_9$ [115]. In their follow-up study, they calculated and compared the ϕ and ψ angle distribution maps of N-acetyl alanine N-methyl amide [116]. Their conclusions from these studies indicated that there were several discrepancies in the performance of the three potentials in terms of achieving the same structure after energy minimization, particularly in cases in which the sample of initial starting configurations is limited. However, they suggested that cases in which the starting ensemble of structures is sufficiently large can converge upon the same configuration.

In a different study, DNA base pairs were used as model systems to test the ability of empirical potentials (including AMBER 3.0, Amber4.1, CHARMM23) and quantum mechanical methods to capture instances of hydrogen bonding [117]. Using several different metrics, they determine that the Amber4.1 potential performs the best because it was developed using 6-31G* restrained electrostatic potential fit charges. Stock and coworkers [118] used trialanine as the model system in which to perform 20 ns MD simulations using a variety of potentials including Amber6 and c22. They calculated several observables including: conformational states and free energies, entropic contributions, and conformational dynamics. They claimed that each force field performed poorly in terms of capturing non-equilibrium dynamical processes.

Though these potentials are most commonly used for simulating biomolecules, they can

also be used in the simulation of small organic molecules. Martin [119] completed a Monte Carlo study to calculate liquid densities and vaporliquid coexistence curves of six organic molecules and compared his results with known experimental values. He used the Amber96 and c22 (along with five other potential functions). In this study, CHARMM outperformed Amber when comparing the 1% error tolerance calculated for liquid properties. In a comprehensive study by Caleman et al. [120], the generalized Amber FF, CHARMM general FF, OPLS-AA and GROMACS were used to calculate the thermodynamics properties including density, enthalpy of vaporization, surface tension, heat capacity (constant volume and pressure), isothermal compressibility, volumetric expansion coefficient, and the static dielectric constant of 146 small organic molecules. They conclude that the OPLS-AA FF best captured the properties of the organic liquids, which comes at no surprise considering OPLS-AA is parameterized for such systems. However, the authors note their surprise at how well Amber and CHARMM do capture the properties of small molecules as those FF were not originally parameterized for those types of systems.

Another important aspect of potential energy functions that can be compared is their ability to compute electrostatic interactions. One recent study [121] uses this as a point of comparison between Amber ff99SB, Amber ff03, Amber ff13α c22, c27, and OPLS 2005. They calculated the formation and breakage of salt bridges between several pairs of residues including: Arg-Asp, Lys-Asp, and His(+)-Asp using 1μ s simulations. They compare their data to experimentally determined K_A values. They conclude that the calculated K_A values are overestimated using each FF. They note that the Amber ff13α force field overestimate the formation of salt bridges the least, while the OPLS 2005 FF overestimates salt bridge formation the most. It should be noted that the used of different solvation models and treatment of the point charges in each model should be taken into consideration for such calculations. They conjecture that using polarizable FF, currently under development, will provide significantly improved treatment of electrostatic interactions in MD simulations.

The ability of a FF to treat the unfolded states of a protein also proves a significant test of the performance of a particular FF. A study was performed by Smith et al. [122] using a 9-residue fragment of the Amyloid-β peptide as a model system and seven popular MD

potentials including c27, Amber99, Amber99SB, Amber99SB-ILDN and Amber03. They determined that metrics such as radius of gyration and solvent accessible surface area were not significantly effected by choice of potential. This suggests that the range of configurations sampled by each FF is approximately similar. However, they identified that some FFs had a propensity to overestimate specific secondary structure motifs and intrapeptide hydrogen bonds differ. For example, they present findings that c27 over-stabilizes helical contacts while Amber99 shows overestimation of intrapeptide hydrogen bonds.

The purpose of the present study is to benchmark the performance of ASMD using c36 [113] and Amber16 [90]. These two FFs are the most recently published FF versions within their respective "family". This work will provide a direct comparison of two popular FFs using two helical peptide motifs: 1PEF and 1LB0. These peptides are used as test systems because they present a significant challenge to both the FF and to the ASMD. This work will add to the existing literature on the comparison of the FFs and can serve as a useful guide to choosing which FF to use in future ASMD studies.

3.1.4 Trends of alanine-rich peptides

Alanine based peptides have been used as model systems to study the helix-coil transition of peptides both experimentally and theoretically. Polyalanine (polyA) repeat segments have been implicated in a number of diseases. It has been previously recognized that polyA segments are the third most common homopeptide repeat in eukaryotic cells behind polyQ and polyN repeats. They are also very common in transcription factors, particularly in repeat segments of 7 residues or more [123]. Larger segments of polyA have been linked to nine human diseases [124, 125]. For example, polyA repeat segments have been linked to be the leading cause of oculopharyngeal muscular dystrophy (OPMD). Though the mechanism of mutation has not been clarified, it is known that the polyA repeats result in misfolding, aggregation, and adverse protein-protein interactions. Therefore, understanding the stability of polyA segments of varying lengths would provide clarity on the overall mechanism by which the segments form.

It has been shown through calorimetry experiments [126, 127] that the $\Delta H_{formation}$

was found to be approximately 1 kcal/mol for the helix-coil transition. Other calorimetry experiments using poly-L-glutamic acid repeats produced the same results shown for the $\Delta H_{formation}$ to be approximately 1 kcal/mol [128]. These calorimetry experiments produced the same results independent of the chain length which shows that this transition is independent of chain length and independent of the residue type.

Computational studies have also been carried out for the helix-coil transition. The model peptides of choice for these studies are polyA homopolymers. Alanine has the highest propensity to form an α-helix [129], which makes it a suitable for studying the helix-coil transition. In addition to it's usefulness as a helix forming residue, the side chain group of alanine contains only a methyl group with four atoms, which reduces the computational cost of the simulations. The limitations of these polyAs is that these polypeptides are insoluble in water.

In vacuum, the stable form of polyAs is an α-helix up to a certain length of peptide (≤ 40), at that point the peptide tends to form a hairpin like structure [130]. There has been much debate over which is the stable conformation of polyAs in solvent, α-helix or random coil. Increasing the length of polyAs in vacuum give rise to more stability of the helix. It has been shown for 12-alanine residues the stable structure is a random coil [131]. Increasing the amount of alanine residues from 12 to 16 increases the stability of the peptide [132]. This has also been shown experimentally as well, increasing the number of alanine residues in a polypeptide increases the helical propensity meaning that the α-helix is the stabilizing structure [133].

In this study we are investigating the additivity of helical peptides. We accomplish this by calculating the PMFs and the hydrogen bonding profiles of the mechanical unfolding of several α-helices with varying lengths of amino acid sequences. Our benchmark systems for this study will be a series of polyAs, ranging from 6 to 50 residues in length. We compare the PMFs, hydrogen bonding profiles, and secondary structure content obtained from the model systems.

3.2 Helical peptide model details

The helical peptides investigated in this Chapter are ALA_{10}, 1PEF, 1LB0 and several polyA peptides of varying lengths and sequences. In the following sections, a detailed description of the design of the systems and simulation parameters used will be presented.

3.2.1 ALA_{10}

The model peptide system ALA_{10} will be used in two different studies in this Chapter. The results of this study comparing the PMFs of ALA_{10} in vacuum, implicit, and explicit solvent, illustrated in Fig. 2, using SMD, ASMD, and FR-ASMD is detailed in Sec. 3.3.1 and has been published [27]. Although this is not the first time that the PMF has been obtained along the stretching coordinate for ALA_{10} in implicit solvent [134], there are two significant advances in the present study: (i) We have obtained the PMF for the vacuum, implicit solvent and explicit solvent cases using a consistent force field for the peptide thereby allowing for a direct comparison of the effects of the varying solvent conditions on the PMF; and (ii) We have obtained the hydrogen-bonding structure along the pathway in each of the three solvent conditions. Below, we compare the free energy functions of unfolding in explicit solvent simulations with the implicit solvent simulations and find qualitative agreement. Near the native state, the free energy profile of the explicit model is slightly more structured. The most significant effect on the PMF is a lowering of the free energy at the unfolded state of circa 2 kcal/mol in the explicit solvent. An additional major finding is that the hydrogen-bonding profiles are qualitatively similar along the pathway between the implicit and explicit solvent simulations.

Throughout this work, we use ALA_{10} to refer to the linear peptide containing ten alanines with the acetylated N-terminus and amidated C-terminus endcaps [23] and which contains only 104 atoms. ALA_{10} lends itself to numerical simulation with low cost. It is notable that there continues to be a debate in the literature over the solubility of these peptides [135, 136, 137]. Short-time numerical equilibration confirms that ALA_{10} in vacuum [23] and in explicit solvent relax to a helical form using the force fields in this work, and provides the starting structures for all of the pulling simulations.

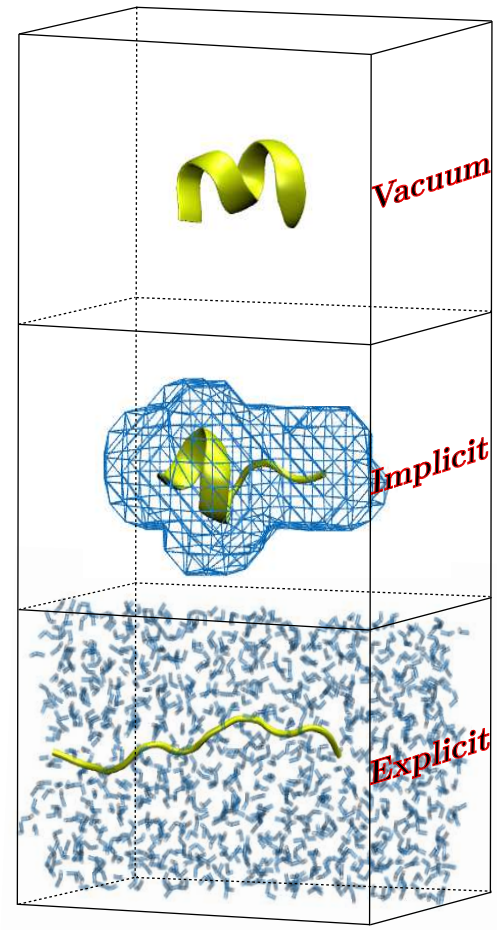

Figure 2: An illustration of the three solvent regimes that are considered for the solvation of ALA_{10} in this work: vacuum (top), implicit solvent (center), and explicit TIP3P water solvent (bottom). In each frame, the ALA_{10} peptide is shown in a different configuration along the pulling coordinate.

For each PMF (such as those shown in Figs. 5, 6, and 7), the reaction coordinate is broken into 10 equal segments, each with 2 Å in length. The overall reaction coordinate corresponds to a total of 20 Å pulling distance, from the folded structure at 13 Å to the fully extended structure at 33 Å. The distance is defined as that between the C_α of the 1st and 10th residues. The pulling force constant, k, is set to 7.2 kcal/mol which is consistent with previous work [23, 24, 55, 56]. The end-to-end distance, measured between the carboxyl nitrogens associated with the two terminal ends, is roughly 13.4 Å. In the propagation of

ALA_{10} using NAMD, the cutoff, switch, and pair list distances are taken to be 12 Å, 10 Å, and 13.5 Å respectively. In all cases, the time step is 2 fs, and temperature is maintained using a Langevin damping coefficient of 5 ps^{-1}.

Equilibration of ALA_{10} for 1 ns in vacuum gives rise to a bound helical structure which is similar to that found by Schulten and coworkers [23] despite the present use of a more recent force field. In explicit solvent, ALA_{10} is solvated in a 20 Å × 20 Å × 50 Å rectangular cuboid with the long-axis arbitrarily labeled as z. The initial positions of ALA_{10} are taken from the vacuum structure with the nitrogen atom of the N-terminus and the nitrogen atom of the cap of the C-terminus placed on the z axis. The box is filled with 897 TIP3P [138] water molecules (as implemented in the psfgen routine associated with NAMD [139]) for a total simulation size of 2795 atoms in accordance with the density of water under these conditions. For the FR-ASMD simulations, 100 ps constrained relaxation stages are used between each pulling stage. Each trajectory is propagated (separately but in parallel) with the end-to-end distance (associated with the nitrogen positions) constrained to their respective initial values. This ensures that no external work is being performed along this (or any) coordinate. For the implicit solvent calculations, the equilibrated vacuum structure was solvated using Generalized Born Implicit Solvent (GBIS) as implemented in NAMD. The cutoff, switch, and pair list distances are also set to 18, 16, and 20, respectively. All propagations are performed using a 2 fs time step and a Langevin damping coefficient of 5 ps^{-1}. The peptide is initially equilibrated before the beginning of the first stage of simulations for 1 ns. As in the explicit solvent, the constrained relaxation stages in the FR-ASMD simulations are propagated for 100 ps.

Results from another study focusing on the dynamical variation obtained using different versions of the CHARMM FF are presented in Sec. 3.3.2. In this work, which is currently in preparation [28], ALA_{10} is also compared to another helical peptide, 1PEF. In addition, the simulation protocols for equilibration and ASMD mechanical unfolding of the homopolymer ALA_{10} across the family of CHARMM potentials remained unchanged from the original publications [27, 56, 55]. Further details on this system are discussed in Sec. 3.2.2.

3.2.2 1PEF

The helical peptide 1PEF is the focus of two studies within this Chapter. First, the mechanical unfolding of 1PEF under different version of the CHARMM FF will be detailed in Sec. 3.3.2. This peptide is also used in the other comparative study using c36 and Amber16 in Sec. 3.3.3 The initial coordinates of 1PEF were obtained from the PDB (PDB code: 1PEF) and the sequence of 1PEF is as follows: EQLLKALEFLLKELLEKL [93]. 1PEF is an 18-amino acid α-helix that is composed of 334 atoms with an acetylated N-terminus and an amidated C-terminus, and all hydrogens are explicitly defined. All MD simulations were performed with NAMD [139] and the c22, c22 with CMAP (c27), and c36 family of potentials. 1PEF was placed along the z-axis of a periodic box composed of approximately 12,600 TIP3P [138] water molecules with the dimensions of 50 Å x 60 Å x 120 Å. The longest side of this square cuboid is chosen so as to ensure that sufficient solvation layers separate the periodic boundary when the peptide is fully extended.

Although the results for ALA_{10} in vacuum, implicit solvent, and explicit solvent using ASMD with the c22 potential were previously published [27, 56, 55], we have reproduced them in the current work using updated scripts. In addition to the reproduced c22 data, the new c27 and c36 ASMD ALA_{10} data results in a robust benchmark for the comparison of the CHARMM family of potentials using our ASMD methodology. This benchmark provides a standard for the hydrophobic, homopolymer helical structure-energetic relationship, and characteristics in each FF. Additionally, the ASMD simulations of 1PEF provided new insight into the complex role of short and long-range noncovalent interactions, which are critical in the folding and unfolding of all macromolecular systems. The simulation protocols for equilibration and ASMD mechanical unfolding of the homopolymer ALA_{10} across the family of CHARMM potentials remained unchanged from the original publications [27, 56, 55]. The following sections detail the simulation protocols for the 1PEF peptide in explicitly represented solvent.

In all equilibrium and ASMD mechanical unfolding simulations of 1PEF, the temperature is regulated using the Langevin thermostat, the van der Waals interaction cutoff distance was set at 12 Å, the smooth switching function beginning at 8 Å, and long-range

electrostatic forces were computed using the particle-mesh Ewald summation method with a grid size of < 1 Å. To prepare the peptide-water system for the ASMD simulations, a minimization and equilibration protocol was followed.

The peptide-water system was initially energy-minimized to remove bad contacts by carrying out a minimization, 10,000 steps of each steepest descent and conjugate gradient method. The system was then slowly heated from 100 - 300 K during a 500 ps simulation (NVT Ensemble) with the solute restrained using 10 kcal/molÅ^2. The system then underwent 1 ns of constant pressure equilibration at 300K (NPT ensemble) through the Nose-Hoover Langevin piston method. A damping coefficient of 5 ps^{-1} with a decay period of 100 fs and a damping time constant of 50 fs was used. The helical peptide's backbone was constrained during the NPT equilibration using a stiff harmonic potential, allowing the water to reach a density of approximately 0.9987 g/cm^3 while allowing the side-chains to reach energetically favorable conformations within the TIP3P water. Next, the peptide-water system underwent a constant volume equilibration (NVT ensemble) at 300K. A series of 100 ps constraint relaxations were performed where the backbone was constrained with a harmonic potential of 10.0 kcal/mol Å^2, 5.0 kcal/mol Å^2, and 1.0 kcal/mol Å^2 consecutively. After the final constraint of 1.0 kcal/mol Å^2 was released, the peptide-water system was allowed to freely equilibrate in the NVT ensemble for 500 ps.

In the production stage of the ASMD simulations, 1PEF is mechanically unfolded such that the end-to-end (r_{ee}) distance between the α-carbon of the N-terminus (ACN) and the α- carbon of the C-terminus (ACC) is gradually increased from 22 to 62 Å. The ACN is kept fixed and the ACC is steered along the longest axis of the periodic square cuboid box, defined to be the z-axis, while harmonically attached to a pseudo particle using a biased Hamiltonian [56]. The harmonic force constant, k, is set to be 7.2 kcal/mol. The reaction coordinate for the forced stretching of 1PEF is defined as the end-to-end distance between the ACN and ACC (r_{ee}). This reaction coordinate was partitioned into 20 stages, such that r_{ee} was perturbed 2 Å/stage. To obtain good statistics, multiple trajectories at each stage have been generated over the equilibrated initial configuration by varying the forces in the Langevin bath. These resulting trajectories are then analyzed numerically and visually with

in-house scripts and by using the VMD package [63].

3.2.3 $\mathbf{gp41}_{659-671}$

The third helical peptide used for studies in this Chapter is $gp41_{659-671}$. Results of the differences and similarities of the energetics, hydrogen bonding patterns, and salt bridge interactions over the course of the forced unfolding of the peptide in c36 and Amber16 are presented in Sec. 3.3.3. The structure of $gp41_{659-671}$ has been studied both experimentally and computationally, although a debate over the stable secondary structure is still on going. The structure of $gp41_{659-671}$ was determined by solution NMR (PDB: 1LCX/1LB0) [140] and is shown in Fig. 3. Biron and coworkers analyzed the structure using various experimental techniques (CD, NOESY, NMR, Sedimentation Equilibrium Measurements, and Diffusion Coefficient Measurements) and concluded that the structure was a monomeric 3_{10}-helix. Another study that performed similar analysis on the peptide concluded that the structure has a propensity for both helix and β-turn [141]. However, a crystal structure was obtained of the 7-mer sequence, $gp41_{662-668}$ ELDKWAS, bound to the antibody 2F5 [142]. The authors found that in the presence of the 2F5, the DKWA sequence forms a β-turn. This finding may have indicated that the secondary structure depends on not only the solvent environment, but on the interaction between the peptide and other biological molecules. Similarly, Crain and coworkers investigated the structure of the peptide in the presence of micelles and how the peptide responds to variations in pH [143]. Interestingly, they found that the peptide only responded to changes in pH in the presence of micelles. When micelles are not present, the peptide remains in an unfolded conformation. There have also been other computational studies performed on the peptide as it's structure presents a challenge for testing computational methods. One quantum mechanical/molecular mechanics study (QM/MM) of 1LB0 concluded it is a 3_{10}-helix, where the peptide is treated using high-level QM methods and the solvent is treated using MM [144]. Several MD studies have been performed, as well. In particular, a MD study utilizing umbrella sampling compares the use of several versions of the Amber and CHARMM FFs [145]. They determined that the peptide has a rugged free energy landscape with shallow minima, allowing it to adopt a

variety of secondary structures. However, the authors were unable to conclude which FF performed the best.

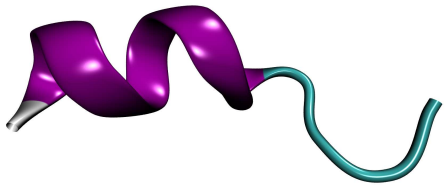

Figure 3: The equilibrated structure of $gp41_{659-671}$ represented as a ribbon. The structure is a mix of α-helix (magenta) and turn (cyan). The side chains have been omitted in this illustration for clarity.

The energetics and structure of the mechanical unfolding of the $gp41_{659-671}$ will be determined using the ASMD method developed in our group and summarized in Chapter 2. The structure of $gp41_{659-671}$ was determined by solution NMR (PDB: 1LCX/1LB0) [140]. This 13-residue fragment is of particular interest because it contains the epitope (sequence of the antigen that is recognized by the immune system) ELDKWA, which causes the response of the monoclonal antibody 2F5 [141]. It exhibits a mix of secondary structure motifs, including α-helix, 3_{10}-helix and turn character, and therefore, presents a significant challenge by which to test the Amber16 and c36 FFs. In order to gain insight into the stability of the secondary structure, the hydrogen bonding profile and specific residue pair interaction strengths will be also monitored and analyzed. Although there have been many computational studies of the peptide [144, 145] and in the presence of other biological molecules [143, 146], there have not been studies that probed the unfolding pathway and stability of the individual peptide sequence.

ASMD simulations will be performed in explicit water solvent with ionization to mimic the natural solvation environment as closely as possible. The pulling velocity of 1 Å/ns will be used to obtain PMFs using ASMD implemented in both AMBER 16 and c36. Additional observables such as hydrogen bonding patterns and residue pair interaction energies will also be calculated.

3.2.4 Alanine-rich peptides

Three different series of peptides will be used in this study to determine the energetic trends and hydrogen bonding profiles of expanding helices. Specifically, ALA homopolymers of lengths 6, 10, 14, 22, 30, 38, and 50 residues will represent the simplest model systems as shown in Fig 4. As there are 3.6 residues per one turn in an α-helix, increasing by the number of residues by factors of 2 adds approximately a half turn. In general, increasing the number of residues in an alanine-rich α-helix increases the stability of the protein. These results will be compared to a water-soluble polypeptide of the general form Ac-Y$(AEAAKA)_k$F-NH2, where k is the number of repeat units from 2 to 8 [127, 147]. This protein increased in α-helical content by increasing the number of (AEAAKA) repeats.

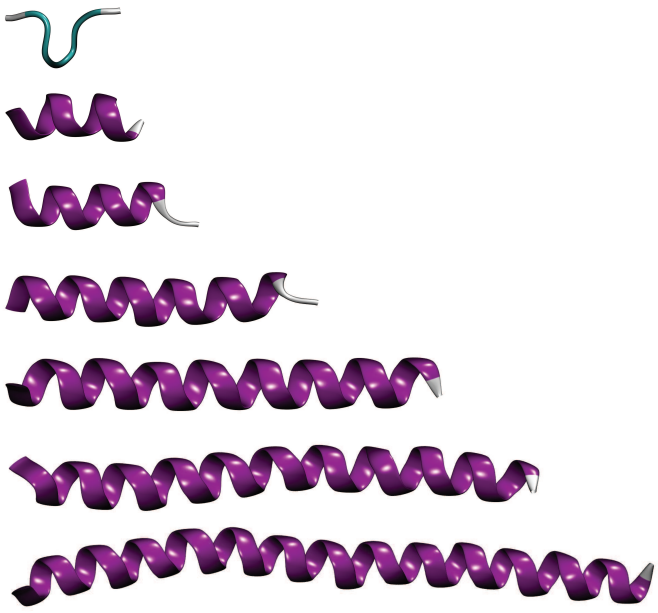

Figure 4: Polyalanines varying with increasing number of residues, 6, 10, 14, 22, 30, 38, and 50, from top to bottom, respectively.

Each structure was built using the VMD plugin molefacture [63]. An acetylated N-terminus and amidated C-terminus were used for the simulation. Each polyA simulation was carried out in explicit solvent conditions. Each structure was placed into a square cubiod solvent box containing TIP3P water molecules. To equilibrate these structures

Table 1: Description of polyalanine systems and ASMD simulation parameters.

peptide	system size (atoms)	length of reaction coordinate (Å)	number of stages
ala6	7329	12	5
ala10	9781	20	10
ala14	12035	28	10
ala22	16081	44	20
ala30	19146	60	25
ala38	24395	76	25
ala50	32225	100	50

the peptides underwent the same three-step equilibration protocol described previously in Sec. 3.2.2. The prevailing structures obtained were α-helices except for ALA_6 which exhibited turn character. The c36 FF [113] and NAMD [139] software were used throughout these simulations. The FF of choice is known be the most accurate of the CHARMM family of potentials and overcomes known issues with previous versions of the FF that favored the formation of α-helical secondary structure.

The length of the unfolded structure is estimated as the number of residues in the peptide multiplied by 2. The ASMD method is used to calculate the PMFs and hydrogen bonding patterns of the polyA peptides. A pulling distance of 2Å per a residue was chosen as the total peptide bond if stretched linear is $\approx$ 3Å. Stretching the protein 2Å per a residue results in an unfolded protein but not necessarily a completely linear, or over stretched, amino acid sequence. The system parameters of each peptide and solvent box is detailed in Table 1. Preliminary results of only the polyA family of peptides are presented in Sec. 3.3.4. This work is currently in progress [30].

3.3 Results and Discussion

3.3.1 Differences between implicit and explicit solvent models

3.3.1.1 Comparison of the energetics of ALA_{10}

The energetics of the directed unfolding of ALA_{10} has been explored using two different variants of SMD: ASMD and FR-ASMD. The simulations are performed at three different pulling velocities, 100 Å/ns, 10 Å/ns, and 1 Å/ns, and in three different environments, vacuum, implicit, and explicit solvent. For each of these pulling velocities and environments,

the number of trajectories was varied from 100 to 800 trajectories per stage so as to achieve convergence in the energetics.

The results presented here include the determination of the PMF using naïve ASMD and FR-ASMD at varying pulling speeds in varying solvent conditions. Specifically, Fig. 5 provides the direct comparison between the converged PMF for ALA_{10} in vacuum, implicit solvent, and explicit solvent. We used a pulling velocity of 1 Å/ns velocity for the vacuum and explicit solvent simulations, which is 10 times slower than our previous work, so as to confirm convergence with respect to the pulling velocity [55]. The convergence of the potentials are illustrated in Figs. 6, and 7, for the explicit and implicit solvents, respectively. The hydrogen-bond profiles along the ASMD pull are shown in Figs. 8, 9 and 10, for ALA_{10} in vacuum, implicit solvent, and explicit solvent. A histogram of the intrapeptide hydrogen bonds as correlated with the actual or effective peptide-solvent hydrogen bonds in each of these three cases is shown in Fig. 11. In the vacuum and implicit solvent cases, the peptide-solvent hydrogen bonds are *effective* in the sense that the relaxation of an all-atom water solvent around the fixed peptide is used to infer the hydrogen bonds to a solvent.

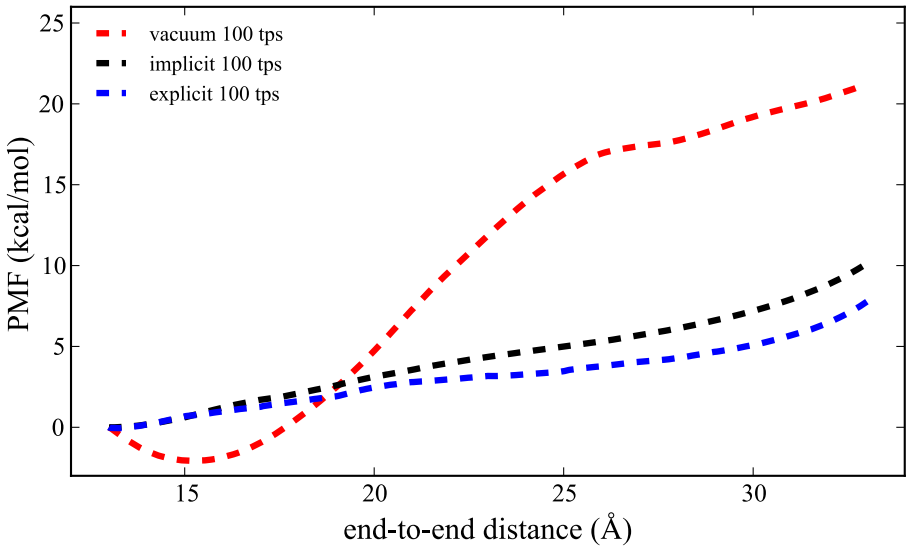

Figure 5: Comparison of the PMF for ALA_{10} in implicit solvent (black curve) to the vacuum (red curve) and explicit solvent (blue curve) results. These PMFs are generated using ASMD at a pulling speed of 1 Å/ns.

A central result of this work is summarized in Fig. 5: the PMF for ALA_{10} in the presence of an implicit solvent is seen to much more closely resemble that of the explicit solvent than the vacuum. Therein, the PMF for the explicit case is considerably lower at the longest extension of the peptide (33 Å), by approximately 12 kcal/mol, than the vacuum case, as was previously reported [56]. Although not shown, the convergence and PMFs studied in the vacuum case were seen to be in near perfect agreement with that reported in earlier work at a velocity of 10 Å/ns [55]. The ΔG from native to unraveled protein in an explicit solvent, as shown in Fig. 6, is 7 kcal/mol which is lower than that found at the faster pulling speeds but the differences have narrowed toward convergence. These results can also be compared to the most converged PMF reported by Tomberli and coworkers [134] using drift-oscillatory steering which found a small lowering in the free energy difference relative to the vacuum result at just below 20 kcal/mol. This suggest that the non-equilibrium ASMD simulations are able to sample pathways with significantly lower energy penalties to the stretching of ALA_{10}.

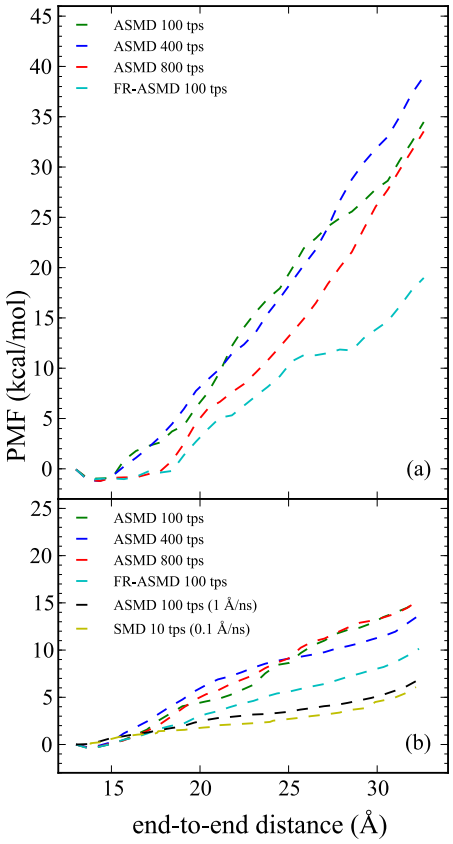

Figure 6: Convergence of the PMF along the ALA_{10} stretching mode in the presence of an explicit solvent calculated using both the ASMD and FR-ASMD methods. The top panel (a) includes the PMFs obtained at 100 Å/ns pulling velocity. The bottom panel (b) includes the PMFs obtained using ASMD and FR-ASMD at 10 Å/ns pulling velocity. In both panels, the red, royal blue, and green curves correspond to ASMD calculations using 800, 400, and 100 trajectories per stage (tps), respectively. A slower pulling ASMD PMF obtained at 1 Å/ns but with only 100 tps is shown in purple in the bottom panel. An even slower SMD simulation obtained a 0.1 Å/ns but with still fewer trajectories (10 tps) is shown in yellow in the bottom panel. In both panels, the light blue curve is the result of the FR-ASMD for 100 tps using 100 ps relaxation stages at the fastest velocity of the ASMD simulations also shown in the corresponding panel.

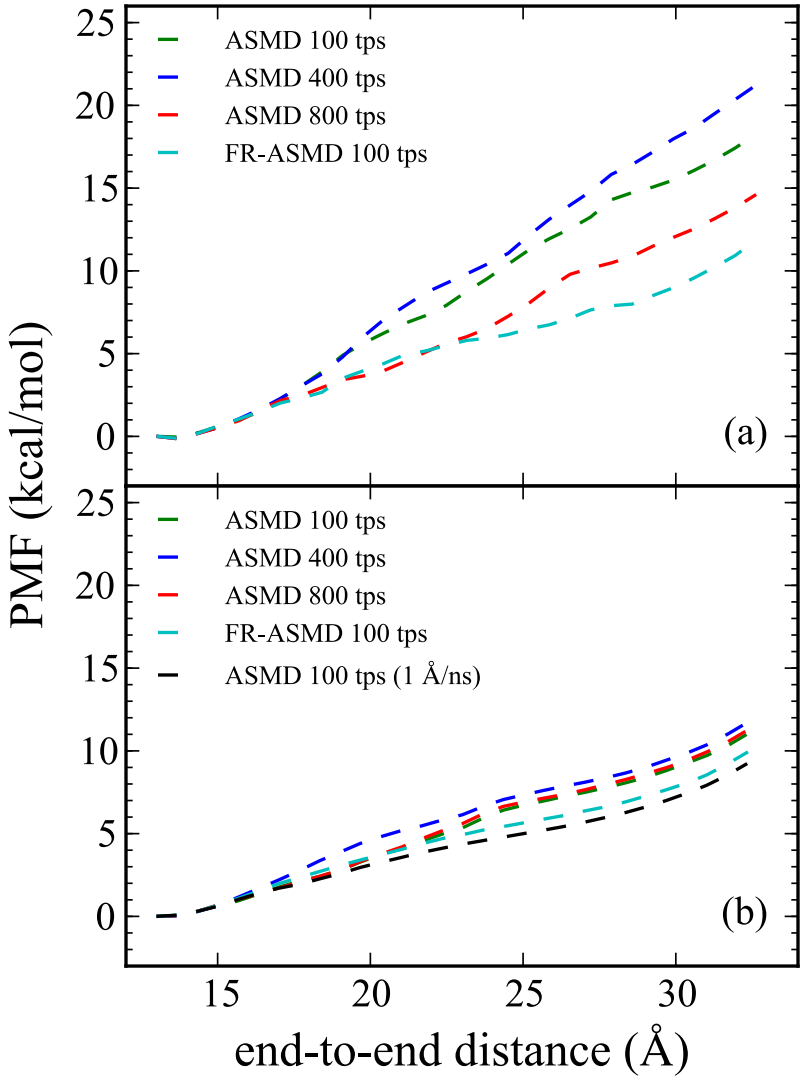

Figure 7: Convergence of the PMF along the ALA_{10} stretching mode in the presence of an implicit solvent calculated using both the ASMD and FR-ASMD methods. All graphs are labeled as in Fig. 6.

Recently, Gumbart and coworkers [70] found a different form for the PMF of ALA_{10} in solvent using adaptive biasing forces and umbrella sampling with the weighted histogram analysis method using the c27 and c36 force fields. In their one-dimensional vacuum PMFs, they found an overall change in the PMF from folded to unfolded in solution to be on the

order of 20 kcal/mol which is comparable to what we and others have found for the same process in vacuum. We find here that the addition of solvent, whether implicit or explicit, gives rise to significant stabilization of the pathway bringing the energy change down to circa 5 kcal/mol as would be expected by the presence of stabilizing hydrogen bonds in the open structure. Gumbart and coworkers also obtained 2-dimensional potentials of mean force for the folding of ALA_{10} in explicit solvent.

They define their reaction coordinates as the end-to-end distance of the peptide (similar to the work we describe here) and as the degree of α-helical content of the peptide. An effective one-dimensional PMF with respect to end-to-end length resulted from the summation over the discrete orthogonal variable. There are significant differences between the 1D-PMF obtained by Gumbart and coworkers and the ones we report in this work. They report a second minimum in the unfolded conformation of the peptide. The differences are likely resulting from (i) sampling, (ii) the choice of reaction coordinate, and (iii) the imposed constraints. In particular, the coordinate orthogonal to the reaction coordinate imposed by Gumbart and coworkers is discrete, but global, and may thereby include structures that are very far (conformationally) to the on-pathway (with respect to the stretching) structures that we are describing. Our non-equilibrium trajectory sampling procedure necessarily suppresses structures that are very far from the pathway which would presumably not be strong contributors in the course of a pulling experiment. These effects likely account for the differences in the free energies that we find along the pathway in comparison with those found in the 2D PMFs by Gumbart and coworkers.

For both implicit and explicit solvent, the decrease of the pulling velocity from 100 to 1 Å/ns leads to convergence of the PMF due to the ability of the peptide to explore more conformational space given more simulation time. The much slower SMD simulation in explicit solvent (0.1 Å/ns) also shown in Fig. 6 is obtained from only 10 trajectories. While this is likely not enough to obtain full convergence, it does provide a rough benchmark for a possible lower bound to the energetics. (It is notable that the PMF for ALA_{10} in vacuum performed at this velocity was converged with only 1 trajectory using SMD [24]). Nevertheless, in explicit solvent there is near agreement between the ASMD (at 1 Å/ns)

and the SMD (at 0.1 Å/ns) suggesting that the former can also serve as a near benchmark for the ASMD simulations. The convergence of the ASMD PMFs in comparison to the near benchmark FR-ASMD result is shown in Fig. 7. The PMF at 10 Å/ns in implicit solvent converges to that at 1 Å/ns more rapidly than in explicit solvent. Thus the lack of explicit solvent fluctuation and relaxation modes in the implicit solvent leads to less spread in the sample space of the non-equilibrium trajectories.

The use of FR-ASMD significantly lowers the PMF achieved by ASMD in both the implicit and explicit solvent cases. FR-ASMD seems to do particularly well at the faster velocity 100 Å/ns in the sense of leading to the most dramatic stabilization in the PMF. (Refer to the light blue curves in Figs. 6 and 7 for the FR-ASMD results.) There are several reasons why the FR-ASMD method decreases the PMF so dramatically. Perhaps the most significant of which is the amount of simulation time that has been added back to the simulation. In the case of the 100 Å/ns pulling simulations, the equilibration time amount to approximately 1 ns, which is comparable to the amount of time required for a corresponding SMD or ASMD simulation obtained at a pulling speed that is ten times slower. It is notable that this discrepancy is not as severe at the slower pulling speeds, when the equilibration time becomes a smaller relative component, but there the gain in accuracy is not as dramatic. Secondly, the contraction to only a single structure from each subsequent stage in ASMD may, in general, also be too stringent for the collapse of the work distributions. Inclusion of the relaxation stages with no contraction to a single structure allows the peptide time to reorient itself with the surrounding solvent molecules, mimicking a slower pulling velocity.

A comparison of the relative computational costs of the ASMD methods is also instructive. All propagations in this work were performed on NAMD in single-core mode on 2.0 GHz AMD Opteron 6128 processors, and reported timings are relative to this resource. The typical CPU time required to propagate a single trajectory (per stage) in explicit solvent at 100 Å/ns and 10 Å/ns pulling is 9 minutes and 93 minutes, respectively. For implicit solvent, the computational time is dramatically reduced to 1 minute and 10 minutes, respectively. FR-ASMD simulations have the same time requirements as in the above cases

for the pulling stages, but require 9 additional constrained relaxation stages (in the present case of 10 pulling stages as used throughout the simulations reported here.) In explicit and implicit solvent, each such constrained relaxations stage required an additional 182 and 24 minutes respectively. This is a substantial penalty, and suggests a need for a more efficient approach for the contraction in ASMD. In summary, naïve ASMD is the most efficient of these, but it requires a check on convergence which may limit its efficiency.

3.3.1.2 Hydrogen bonding profiles

The hydrogen bonding profiles, in conjunction with the PMFs, reveal interesting structural characteristics of the unfolding of ALA_{10}. As previously reported [55], the change in the PMF between the end points for ALA_{10} in vacuum is higher than that found in solvent partly because of differences in the hydrogen bonding structure. During the unfolding of ALA_{10} in vacuum, the number of $i \rightarrow i+4$ (α-helix) contacts, where i refers to the index of the residue, decreases steadily as shown in Fig. 8 after initially being insensitive to the extension. The loss of $i \rightarrow i+4$ contacts appears to be coupled with the onset of $i \rightarrow i+3$ (3_{10}-helix) contacts just before an end-to-end distance of 18 Å. The number of $i \rightarrow i+3$ contacts reach a maximum at an extension of 23 Å, and are thereafter precipitously lost. There is no formation of $i \rightarrow i+5$ (π-helix) contacts. The exchange of $i \rightarrow i+4$ contacts with $i \rightarrow i+3$ contacts was seen earlier to lead a near constant total number of hydrogen bonds for the first 15 Å in extension between the ends of the peptide.

This behavior is in stark contrast with the hydrogen-bonding profile observed in explicit solvent shown in Fig. 10. In explicit solvent, there is no formation of the $i \rightarrow i+3$ contacts while the $i \rightarrow i+4$ contacts are lost [56]. This is due to the presence of water molecules that are able to hydrogen bond with the peptide as it is stretched. Unlike in the vacuum case, there is slight formation of the $i \rightarrow i+5$ (π-helix) contacts. This finding is corroborated by a similar conclusion about the role of hydrogen bonding in human amylin seen earlier by Skinner and coworkers [66]. Namely, that hydrogen bonds from the solvent significantly contribute to the stabilization of the unfolded helix in random coil.

As shown in Fig. 9, we find only small differences in the hydrogen-bonding trend between

the implicit and explicit solvent cases in accord with the good agreement seen earlier between their PMFs. In both cases, the trend is also unlike that seen in the vacuum, but this is to be expected as the solvent provides sufficient solvation to limit the dramatic formation of $i \to i+3$ contacts seen in the vacuum case. The main difference, albeit small, between the implicit and explicit solvent hydrogen-bonding profiles appears to be due to the fact that the initial equilibrium structures are quite different. Namely, ALA_{10} in implicit solvent appears to have significant π-helical structure as observed from the presence of a larger number of $i \to i+5$ contacts comparable to the number of $i \to i+4$ contacts. The loss of π-helical structure with increased end-to-end distance initially results in α-helical structure rather than no structure at all.

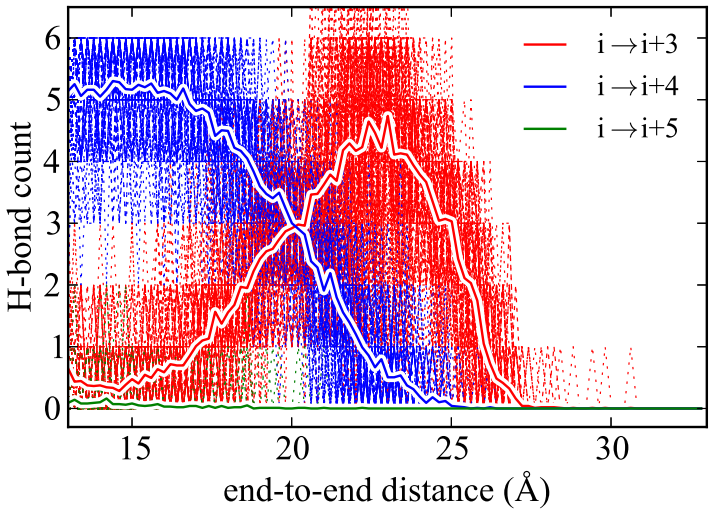

Figure 8: Hydrogen bond profile for ALA_{10} in vacuum calculated using ASMD at simulation conditions, 1 Å/ns with 100 tps, differing with those of our earlier work [55]. The red, blue and green solid lines corresponds to the average number of $i \to i+3$ (3_{10}-helix), $i \to i+4$ (α-helix), and $i \to i+5$ (α-helix) contacts, respectively. They are overlaid onto scatter plots showing the range of values visited by the specific trajectories, thereby indicating the error in the averages.

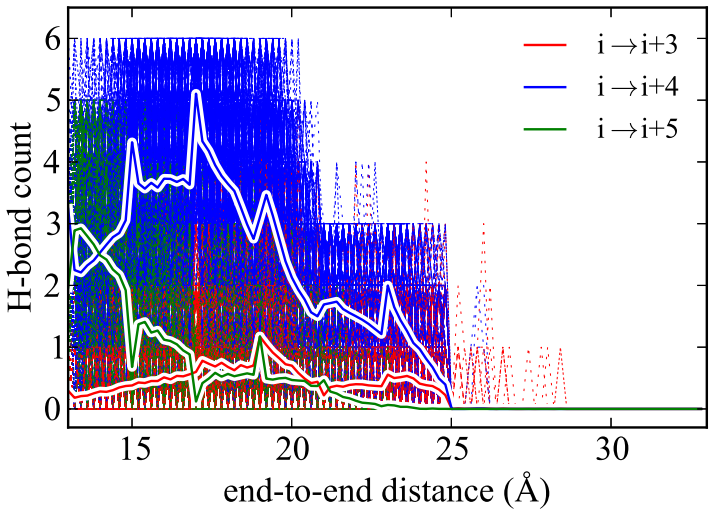

Figure 9: Hydrogen bond profile of ALA_{10} in implicit solvent using ASMD at 10 Å/ns with 800 tps. All curves are shown as in Fig. 8.

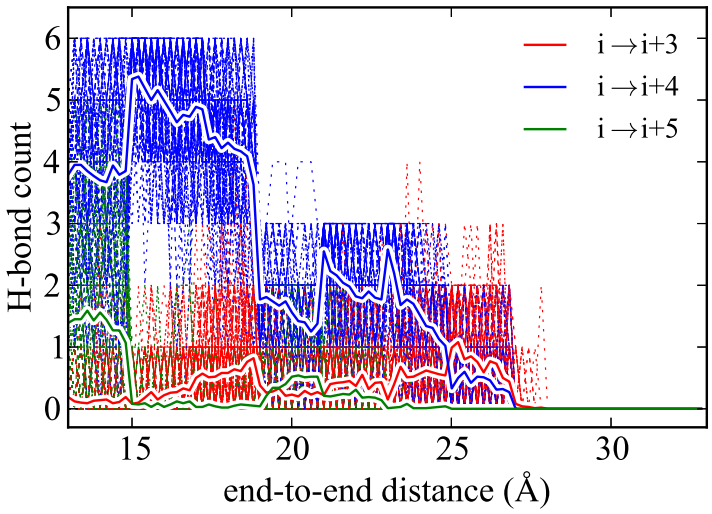

Figure 10: Hydrogen bond profile for ALA_{10} using ASMD at 10 Å/ns with 100 tps recalculated here for the explicit solvent case [56]. All curves are shown as in Fig. 8 and 9.

The role of hydrogen bonds between the peptide and water can be monitored readily in an explicit solvent by enumerating the contacts along the pathway. Fig. 11 (top) displays a histogram of the peptide-peptide and peptide-solvent hydrogen bonds (for one thousand

structures) along the most favored pathway (as determined by the JE-criterion) as the peptide is stretched. Unfortunately, such a procedure is not readily available in an implicit solvent (and certainly not in vacuum), making it difficult to confirm the hypothesis that the hydrogen-bonding to solvent is somehow being characterized appropriately by the implicit solvent as suggested by the analysis above.

Instead, we obtain the effective number of peptide-solvent hydrogen bonds that would be created if the solvent were to be equilibrated around the fixed peptide structures. The histograms for the peptide-peptide and effective peptide-solvent hydrogen bonds along the most favored pathway (as determined by the JE-criterion) as the peptide is stretched in the implicit solvent and vacuum are shown in Figs. 11 (middle) and 11(bottom), respectively.

The similarity between the histograms for explicit (top) and implicit solvent (middle) in Fig. 11 is startling. In all three cases, the number of intrapeptide hydrogen bonds decreases steadily as peptide-solvent hydrogen bonds increase. However, the relative abundance is strongly shifted toward the case when many peptide-solvent hydrogen bonds are favored only in the cases of the implicit and explicit solvation. In the vacuum, the largest abundance of hydrogen bonds occurs in the regime when they are primarily intramolecular. That is, the opening of the peptide is not followed by structures that would be readily solvated through hydrogen-bond contacts. This should not be surprising for the vacuum case because the structures are formed in absence of water. However, the fact that the implicit solvent induces structures that readily admit hydrogen bonds in near proportion as those seen in the explicit solvent case further suggests that the former does indeed include hydrogen-bonding (implicitly) at a level of detail that might not have been anticipated.

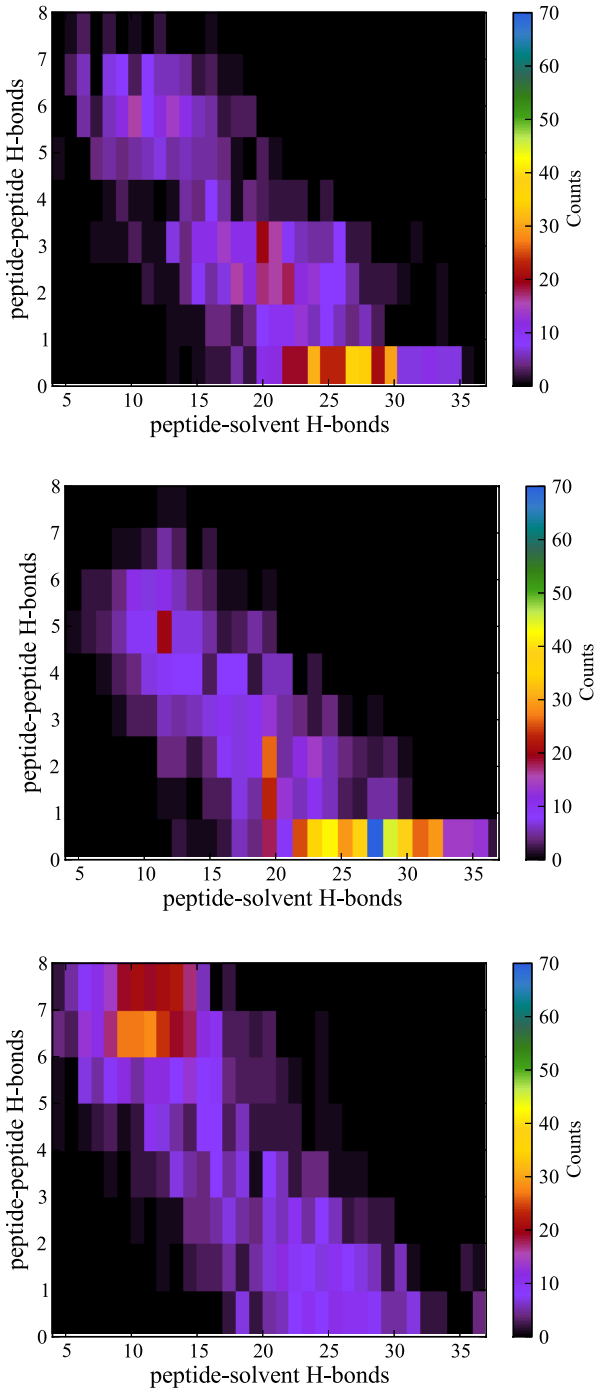

Figure 11: Two-dimensional histogram of peptide-peptide and peptide-solvent hydrogen bonds for ALA_{10} structures in explicit, implicit, and vacuum along an ASMD pulling path. The color scale on the right from black to blue indicates the density of the points for that region of the map.

Thus, the overall pathway seen in the implicit solvent appears to involve the loss of internal helical structure that is predominantly replaced by solvent contacts in agreement with that observed in the case of explicit solvent. This method of solvating and equilibrating the implicit and vacuum structures is simply used as a tool to understand how the structures respond to solvation.

3.3.2 Benchmarking the energetics across different CHARMM potentials

3.3.2.1 Structural Analysis

Each of the CHARMM force fields—c22, c27 and c36—in this study give rise to very different weightings in the structures of the corresponding equilibrium ensemble. These differences can be characterized by comparison of the representative structures of the ensembles and the minimum energy structures found in the ASMD-generated PMFs. These structures also differ from the corresponding structures in the PDB because the latter are generally obtained under different conditions, and there may also be systematic error due to the differences in the force fields. Nevertheless, in order to ensure consistency and reproducibility, protein structures from the literature serve to initiate all of the structural determinations in the numerical experiments described here.

Equilibrated Structures: $\mathbf{ALA_{10}}$ and 1PEF The two α-helical peptides used in this study, ALA_{10} and 1PEF, were minimized and equilibrated as outlined in Sec. 3.2.2. The initial structure for the α-helical peptide ALA_{10} is specified by the set of coordinates of the 104-atom model employed in earlier SMD studies by Park and Schulten [23, 24]. The initial structure for 1PEF is taken from X-ray-diffraction (XRD) data readily available in the PDB [93]. As discussed in Sec. 3.2.2, once the system has reached and maintained equilibrium, we choose the last structure in the trajectory as the representative equilibrated structure of the corresponding ensemble. (We confirmed that initialization of our procedure using other structures selected from the equilibrated trajectory led to negligible differences in the observables discussed here.) The end-to-end distances $r_{\mathrm{ee}_{\mathrm{equil}}}$ of the equilibrium structures for ALA_{10} and 1PEF as modeled by the three CHARMM forces fields under investigation in this study are reported in Table 2. The structures were then contracted and pulled

t

Table 2: The end-to-end distances ($r_{ee_{equil}}$) of the equilibrated α-helices ALA_{10} and 1PEF for each CHARMM potential are compared to those ($r_{ee_{min}}$) of the minimum energy structures obtained using our ASMD methodology across the family of CHARMM potentials. The previously reported values of $r_{ee_{min}}$ for ALA_{10} obtained using SMD or ASMD are also listed with the corresponding reference number in parent. Note that the PMFs for ALA_{10} modeled using c22 exhibited no minimum in the observed region and this is indicated by the dash.

	CHARMM FF	$r_{ee_{equil}}$	$r_{ee_{min}}$ **(ASMD)**	$r_{ee_{min}}$**(Ref. #)**
ALA_{10}	c22	12.6 Å	–	– ([27])
	c27	14.1 Å	14.3 Å	14.3 Å([56])
	c36	12.7 Å	14.4 Å	14.3 Å([70])
1PEF	c22	25.8 Å	24.3 Å	
	c27	24.4 Å	25.5 Å	
	c36	25.9 Å	25.8 Å	

according to the ASMD protocol with the resulting minimum in the PMF found at $r_{ee_{min}}$ as also listed in Table 2. In the case of ALA_{10}, these latter values compare well to those previously obtained using either SMD or ASMD.

Comparisons between the resulting equilibrated peptide conformations for ALA_{10} and 1PEF using the three selected CHARMM FFs are shown in Fig. 12. The equilibrated peptide conformation for each FF is cross-compared with the others of that peptide — *i.e.*, the 1PEF c22 equilibrated structure (red) was aligned with the equilibrated c27 initial ASMD structure (green)— to obtain their relative RMSD. The computation of the RMSD values includes side chains because the comparison between the three force fields necessarily hinges on the differences in how each treats the side-chain torsions. The resulting deviations in the RMSDs vary more widely with FF for the ALA_{10} peptide than for 1PEF. The RMSD difference between the c22 and c27 structures of ALA_{10} was the smallest observed at 1.33 Å, whereas the difference between the c27 and c36 structures was the largest at 2.88 Å.

The trends in the RSMD values for ALA_{10} shown in Fig. 12 appear to echo the known incremental changes across the three force fields. Recall that the primary differences between the c22, c27 and c36 FFs are the inclusion of the CMAP grid correction to c22 accounting for the π-helical artifact of the c22 FF, and a reparameterization in c36 to address α-helical bias in c27. In contrast to the large differences in the RMSD values seen for the

model system ALA_{10}, the peptide 1PEF resulted in RMSD values across the family of FFs that are approximately equivalent; 2.45 Å, 2.35 Å, and 2.72 Å respectively. This is somewhat surprising because 1PEF is helical and hence should be affected by the changes in the terms of each force field. However, 1PEF is a complex protein with numerous intra-peptide interactions that affect the α-helical peptide stability beyond the simple geometric terms related only to the helical turns. This suggests that the structure of 1PEF may be primarily stabilized by interactions unrelated to those from CMAP. This could include interactions between charged side chains through formation of salt bridges, and solvent-mediated interactions, such as hydrogen bonding (Sec. 3.3.2.4), result. The resolution of this additional puzzle is being investigated in on-going work.

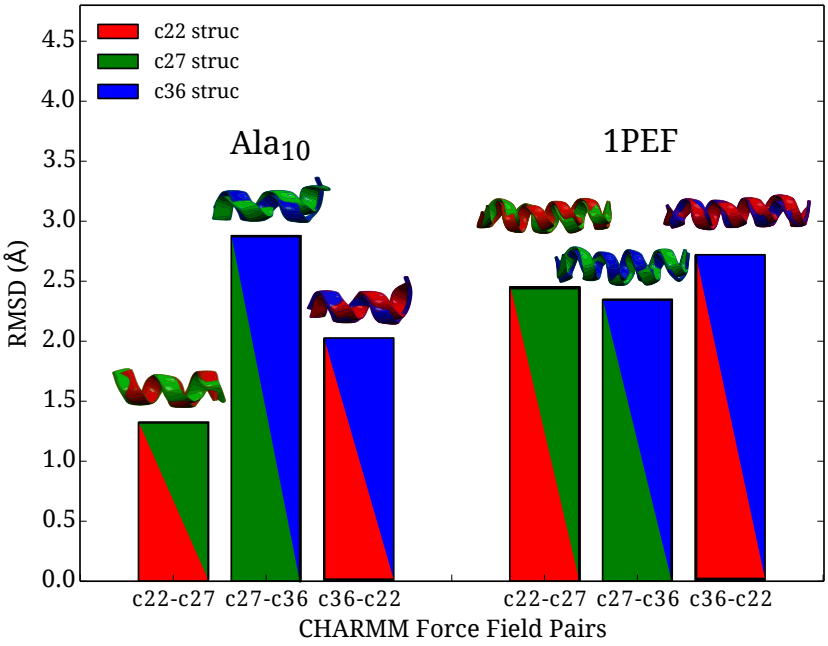

Figure 12: Structures and RMSDs between the initial equilibrated explicit solvent structures of ALA_{10} and 1PEF obtained from the c22 (red), c27 (green) and c36 (blue) FFs.

Minimum Energy Structures: 1PEF According to the published XRD structure [93], 1PEF has significant helical character and contains critical intramolecular contacts, while

exhibiting some helical deformations. Utilizing the PMF generated for each ASMD simulation (in particular, the 100 tps at 1 Å/ns in each CHARMM FF) such as those shown in Fig. 14, the minimum energy structure was determined. For each non-equilibrium trajectory, there exists a minimum energy structure which can be assigned according to the one whose extension matches that at which the PMF has a minimum. An average minimum energy structure ($\overline{X}$) was obtained by averaging 100 minimum energy structures using custom scripts and VMD. The XRD 1PEF structure is illustrated as a purple ribbon in Fig. 13. It is compared through backbone alignment to the average minimum energy c22 structure ($\overline{X}_{c22}$, red ribbon), the average minimum energy c27 structure ($\overline{X}_{c27}$, green ribbon), and the average minimum energy c36 structure ($\overline{X}_{c36}$, blue ribbon). These 1PEF backbone alignments form the basis for the metrics listed in Table 3. A similar comparison for ALA_{10} is not shown because all the ASMD simulations for this peptide have been initiated with the same structure as that initially used by Park and Schulten [23, 24]. It has been inserted into an explicit water box as outlined in the Methods. The ALA_{10} minimum energy structures do differ across the selected CHARMM FFs as indicated by the differing values of the corresponding r_{ee} distance listed in Table 2.

The end-to-end distance (r_{ee}), the RMSD when compared to the XRD structure, and the average ϕ and ψ angles along the backbone of 1PEF for the experimental and computed peptide structures are listed in Table 3. The character of the possible backbone (not side-chain) intrapeptide hydrogen bond contacts for the experimental minimum energy structures of 1PEF are provided in Table 4. The ψ and ϕ angles in the XRD structure of 1PEF in residues 2 through 15 consistent with α-helical character as indicated by the average values of those angles, $\bar{\phi}(2, 15)$ and $\bar{\psi}(2, 15)$ in Table 3, and more precisely by each pair of angles (not shown.) Also typical of α-helices is the observation that 12 of the 14 main chain intramolecular hydrogen bonds, between residues 1 and 16, are $i \rightarrow i + 4$ contacts, as indicated in Table 4. 1PEF exhibits a distortion away from ideal α-helical structure in residues 16 through 18. One pair of ϕ_i and ψ_i angles at $i = 16$ has a slightly distorted stereochemistry, while the second pair at $i = 17$ is more consistent with the typical range of angles observed for 3_{10}- helical conformations. Similarly, the relative poses of the final

residues 16 to 18 give rise to $i \rightarrow i+3$ hydrogen bond contacts, thus exhibiting 3_{10}-helical character. Helical peptides exhibiting structures in which α-helical character dominates in the middle of the stable helix while 3_{10}-helix character dominates near the C-terminus, as in 1PEF, have been observed by several groups [148, 149, 150, 151].

The secondary structural elements of the experimental 1PEF peptide are generally captured by each CHARMM FF in the average minimum energy structure, however there are some discrepancies. The $\overline{X}_{c27}$ and $\overline{X}_{c36}$ structures contain only α-helical backbone intrapeptide hydrogen bonds, as indicated in Table 4. This finding is further supported by the results in Sec. 3.3.2.3 on the secondary structure evolution during the mechanical unfolding event, and those in Sec. 3.3.2.4 on the hydrogen bonds explicitly accounted for along the forced unfolding pathway. However, the $\overline{X}_{c36}$ structure only differs by a RSMD of 0.58 Å from the experimental structure, and the ψ_{16} and ψ_{17} torsional angles are closer to the XRD structure. On the other hand, the average 1PEF structure from the c22 ASMD simulations ($\overline{X}_{c22}$) exhibits some unusual structural features. The r_{ee} distance is approximately 2.5 Å shorter than the experimental structure, in addition to the ψ_{16} and ψ_{17} torsional angles being overestimated (Table 1). Furthermore, the $\overline{X}_{c22}$ structure has an RMSD difference from the XRD structure by over 2 Å. The decrease in r_{ee} distance of the $\overline{X}_{c22}$ 1PEF peptide is a consequence of two factors: (*i*) a single 3_{10}-helical contact ($i \rightarrow i+3$.) occurring between Ala6 and Phe9, which causes the helix to bend, and (*ii*) a π-helical contact ($i \rightarrow i+5$) between residues Glu13 and Leu18 which increases the pitch or width of the helix (Table 4). The formation of the π-helical contact in the minimum energy structure of 1PEF in the c22 FF is expected, and is further supported by our hydrogen bonding data (see Fig. 17 in Sec. 3.3.2.4) and our secondary structure analysis (see Fig. 16 in Sec. 3.3.2.3).

Table 3: Comparison between the XRD 1PEF structure and the average minimum energy structures ($\overline{X}_{FF}$) shown in Fig. 13 calculated for each ASMD simulation with $v_s = 1$ Å/ns and 100tps. Values displayed include: the end-to-end distance $r_{\mathrm{ee}_{\mathrm{min}}}$ (in Å) of the averaged minimum energy structure, the RMSD (in Å) between the experimentally determined 1PEF coordinates and minimum energy structures, and the ϕ and ψ backbone torsional angles. The specific angles for residues 2 through 15 are nearly the same and hence only the averages, $\bar{\theta}(i,j) \equiv (j-i+1)^{-1} \sum_{k=i}^{k} \theta_k$, and corresponding RMSDs are shown for θ being ϕ and ψ, $i = 2$ and $j = 15$.

	$r_{ee_{min}}$	**RMSD**	$\bar{\phi}(2,15)$	$\bar{\psi}(2,15)$	ϕ_{16}	ψ_{16}	ϕ_{17}	ψ_{17}
Exptl.[93]	26.9 Å	-	-62±3°	-42±4°	-87°	-4°	-63°	-16°
$\overline{X}_{c22}$	24.3 Å	2.02 Å	-64±9°	-43±6°	-71°	-57°	-84°	-53°
$\overline{X}_{c27}$	25.5 Å	0.86 Å	-63±1°	-41±3°	-65°	-45°	-84°	-26°
$\overline{X}_{c36}$	25.8 Å	0.58 Å	-63±3°	-40±3°	-66°	-38°	-74°	-30°

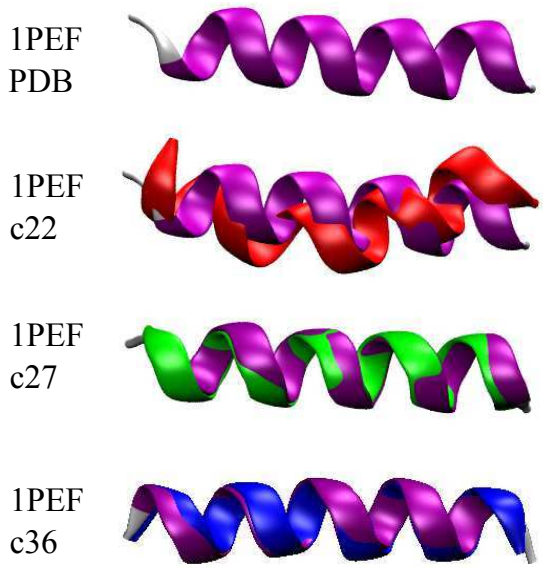

Figure 13: Comparison between the XRD 1PEF peptide (purple) and the average minimum energy structures—$\overline{X}_{c22}$ (red), $\overline{X}_{c27}$ (green) and $\overline{X}_{c36}$ (blue)—obtained using ASMD simulations with $v_s = 1$ Å/ns and 100tps.

3.3.2.2 Energetics

The energetics along the forced unfolding pathway of ALA_{10} and 1PEF in an explicit (water) solvent has also been obtained using ASMD at two pulling velocities v_s, 10 and 1 Å/ns, with 100 trajectories per stage (tps). The ASMD simulations for ALA_{10} are performed in

Table 4: All intrapeptide hydrogen bond contacts for the experimental structure of 1PEF, and the minimum energy structures of 1PEF in each CHARMM FF. π-helical contacts are $i \rightarrow i+5$, α-helical contacts are $i \rightarrow i+4$, and 3_{10}-helical contacts are $i \rightarrow i+3$.

Donor	**Acceptor**	**Exptl.[93]**	$\overline{X}_{c22}$	$\overline{X}_{c27}$	$\overline{X}_{c36}$
LEU18	GLU13	-	π	-	-
LEU18	LEU15	3_{10}	-	-	-
LEU18	LEU14	-	-	α	α
LYS17	LEU14	3_{10}	-	-	-
LYS17	GLU13	-	α	α	α
GLU16	GLU13	3_{10}	-	-	-
GLU16	LYS12	α	α	α	α
LEU15	LEU11	α	α	α	α
LEU14	LEU10	α	α	α	α
GLU13	PHE9	α	-	α	α
LYS12	GLU8	α	-	α	α
LEU11	LEU7	α	α	α	α
LEU10	ALA6	α	α	α	α
PHE9	LYS5	α	α	α	α
PHE9	ALA6	-	3_{10}	-	-
GLU8	LEU4	α	α	α	α
LEU7	LEU3	α	α	α	α
ALA6	GLN2	α	α	α	α
LYS5	GLU1	α	-	α	α
	Numbers of Contacts:	12: α 3: 3_{10}	10: α 1: 3_{10} 1: π	14: α	14: α

10 equally-spaced incremental stages covering a change in the overall pulling coordinate of 20 Å. The ALA_{10} peptide is pulled from the initially compressed α-helical structure with an r_{ee} distance of 13 Å to a fully extended structure with a final r_{ee} distance of 33 Å. Twenty equally-spaced incremental stages are employed for the ASMD simulations of 1PEF covering a change in the overall pulling coordinate of 40 Å. The 1PEF peptide is pulled from the initially compressed α-helical structure with r_{ee} distance of 22 Å to a fully extended structure with a final r_{ee} distance of 62 Å. The PMFs obtained by ASMD are shown in Fig. 14. Several pulling velocities have been employed because earlier work on SMD [152, 23, 22, 24] and ASMD [55, 56] indicated that it can affect the overall energetics and pathways. The slowest pulling velocity employed here ($v_s = 1\,$Å/ns) was seen to be sufficient for confirming convergence in our previous work on the ALA_{10} peptide[55, 56, 27] and two β-turn peptides.[31] It appears to be sufficient here as the insets in Fig. 14 displaying

the first 10Å of the pull obtained at 1 Å/ns are in agreement with the 10 Å/ns pulls.

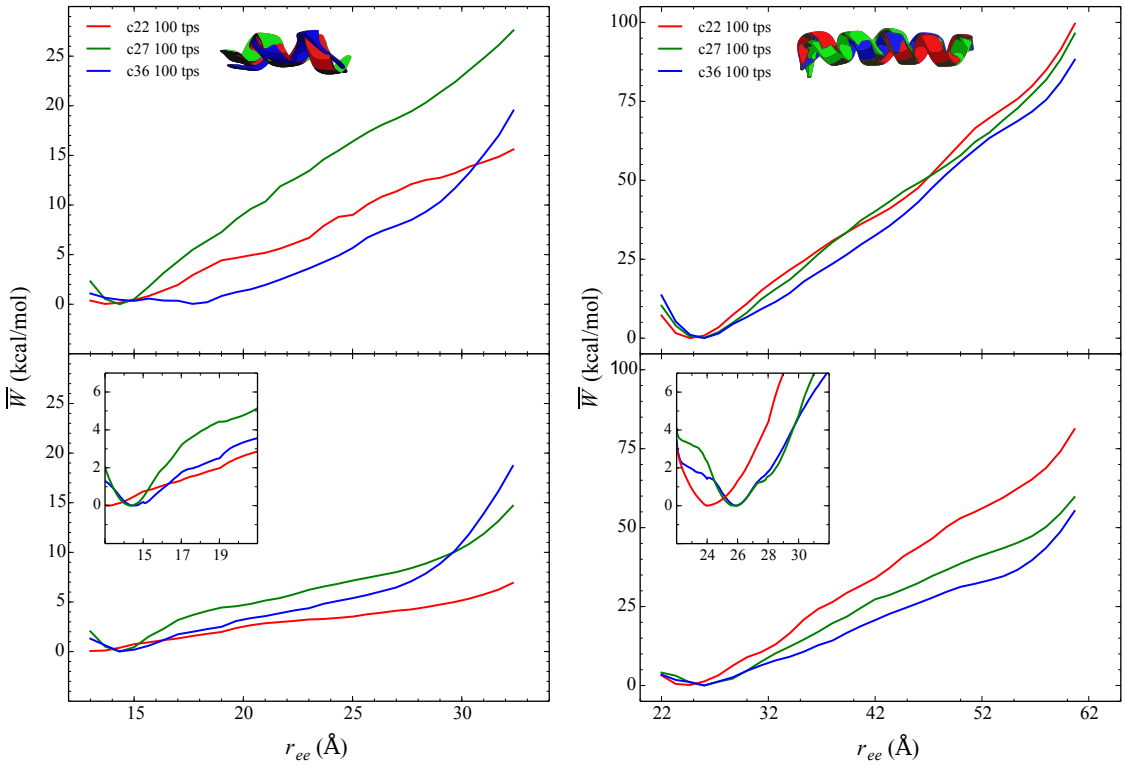

Figure 14: The comparison of the explicit PMFs obtained using the CHARMM family of potentials for the forced stretching of ALA_{10} (left) and 1PEF (right) using ASMD, as specified in the legend. The PMFs are displayed for two stretching velocities, 10 Å/ns (top panels) and 1 Å/ns (bottom panels), and are obtained by averaging over 100 trajectories per stage (tps). The inset in the bottom panel displays the behavior of the PMFs obtained at a slower 1 Å/ns velocity near the respective peptide's energy minimum structure.

The structures of ALA_{10} and 1PEF at the start of the ASMD pulling simulations are not the minimum energy structures as can be seen from the PMFs in Fig. 14 for both pulling velocities. The initial structures for ALA_{10} and 1PEF are therefore seemingly arbitrary in the sense that they do not correspond to a set point such as the minimum of the PMF. This is a consequence of the fact that the peptides are intentionally made more compact, following the original procedure of Park and Schulten [23, 24], so as to allow SMD (and

ASMD) to find the minimum through the pulling procedure. Meanwhile, the extension $r_{\mathrm{ee}_{\mathrm{min}}}$ (ASMD) of the peptide at which the ASMD PMF is a minimum is not quite equal to the typical extension $r_{\mathrm{ee}_{\mathrm{equil}}}$ at which the peptide is freely equilibrated as can be seen in Table 2. The discrepancies likely arise from the relative numerical error in the two methods. It is also possible that the SMD or ASMD procedure could limit the sampling of structures in such a way as to bias the non-equilibrium average. To mitigate and expose these effects, we have used slow pulling speeds while displaying comparisons to the results obtained from faster pulling speeds. The fact that convergence is possible has already been seen in the case of ALA_{10} [23, 24, 55].

For both steered unfolding velocity simulations of ALA_{10} (left panels), c22 (illustrated as the red curve) does not have a free energy minimum, whereas c27 (green curve) and c36 (blue curve) both have a clear energy minimum. In particular, for the slow velocity steering simulation of ALA_{10} (bottom left panel), the global minimum of the potential energy curve has a corresponding well depth of 2.03 kcal/mol, with an r_{ee} of 14.3 Å and 1.30 kcal/mol, with an r_{ee} of 14.4 Å respectively. The minimum energy structures of ALA_{10} are a well defined α-helix. Even though the c22 PMF does not display a minimum energy structure, the structural features are mostly α-helical with a small percentage of π-helical characteristics. The presence of π-helical contacts are not an unexpected trend for the c22 FF [103]. The c27 and c36 minimum energy conformations of ALA_{10} have approximately the same r_{ee} value, 14.3 Å and 14.4 Å respectively, and very similar free energy values. However, the c27 ALA_{10} conformation is lower in energy by 0.73 kcal/mol. Both of these minimum energy structures posses only α-helical contacts, with the c27 structure having 2 more α-helical contacts than the c36 structure. The over stabilization of the α-helix ALA_{10} in the c27 FF, through an increase in the $i \rightarrow i+4$ hydrogen bond contact count, explains the lowering in energy of the c27 ALA_{10} structure.

However, unlike ALA_{10} in the slow velocity steered simulations, the c22 simulation exhibits a minimum energy structure for 1PEF with an r_{ee} of 24.30 Å with a well depth of 3.2 kcal/mol. When comparing the c22 minimum energy structure to the structures from the c27 and c36 simulations, the c22 minimum energy structure is shifted to shorter r_{ee}. As

previously discussed in Sec. 3.3.2.1, the c22 minimum energy structure is over 2.5 Å shorter from end-to-end than the experimental XRD structure, and approximately 1.5 Å shorter from end-to-end than the c27 and c26 minimum energy structures. The shift to shorter r_{ee} in c22 is resultant from a terminal π-helical contact, defined as $i \rightarrow i+5$, making the helix shorter and fatter, in addition to a centrally located 3_{10}- helix contact causing the c22 1PEF structure to bend. The c27 and c36 1PEF minimum energy structure follow the same trend as the model system ALA_{10}. The two structures have very similar r_{ee} values (25.50 Å and 25.80 Å respectively), and differ in free energy by 0.60 kcal/mol (with well depths of 4.1 kcal/mol and 3.5 kcal/mol respectively). Both of these minimum energy structures posses α-helical contacts, but unlike the model system, the c27 and c36 1PEF minimum energy structures have the same amount and type of hydrogen bonds, 14 α-helical $i \rightarrow i+4$ contacts. There is an over stabilization of the c27 1PEF α-helix, but it is not a result of an increase in the number of $i \rightarrow i+4$ hydrogen bond contact counts. This 1PEF c27 result more clearly illustrates the reported α-helical bias of the FF, and suggests that short range interactions do not fully account for their stability as remarked earlier.

There is a drop in the overall magnitude of the PMFs for both the ALA_{10} and 1PEF peptide systems as the v_s is decreased from 10 Å/ns to 1 Å/ns. The drop in the overall PMF for ALA_{10} was approximately 10 kcal/mol regardless of which of the FFs was used in the ASMD simulations. The corresponding drop for 1PEF is approximately 20 kcal/mol. This is consistent with the fact that the latter has a helix that is twice and that the stabilization energy is approximately additive. Across the CHARMM family of potentials for the slow v_s extension of ALA_{10}, the thermodynamically accessible states remain fairly structured in the sense that there is an initial fast rise from the energy minimum conformation, and then a subsequent flattening of the PMF up to $\approx$ 23 Å r_{ee}, where the peptide becomes completely unstructured. The homopolymer ALA_{10} is primarily stabilized through backbone hydrogen bond contacts, as there are no charged residues for longer ranged electrostatic interactions, such as through salt bridge formation. The secondary structure analysis shown in Fig. 15 and the hydrogen bond analysis in Fig. 17 clearly supports this conclusion. Moreover, ALA_{10} (leftmost column) has little to no remaining secondary structural characteristics

after approximately the first 10 Å of extension. On the other hand, the lack of a flattening of the PMFs for 1PEF suggest that it remains highly structured through along the induced unfolding pathway by way of retention of secondary structural elements. The 1PEF PMFs across each CHARMM potential have a fairly consistent positive slope for both v_s. A positive and consistent slope across all FFs, for both v_s, demonstrates a positive correlation between the energetics and r_{ee}, which directly correlates to the observed secondary structure along 1PEF's induced unfolding pathway.

3.3.2.3 Secondary Structure Evolution

In the following, we have compared the secondary structure alterations as a function of peptide extension from our ASMD simulations of ALA_{10} and 1PEF. The STRIDE algorithm from VMD in conjunction with customs scripts was used to generate Figs. 15 and 16, in which the ALA_{10} results and the 1PEF results are displayed. In particular, the ASMD explicit 1 Å/ns trajectory with 100 tps that was nearest to the computed JA was chosen for this analysis. The c22 FF results are shown in the top panels, the c27 FF results are the middle panels, and the c36 FF results are in the bottom panels. The evolution of the change in secondary structure as a function of mechanical unfolding is displayed as: α-helical (purple), π-helical (red), 3_{10}-helical (blue), turn (teal), and random coil (black in leftmost columns, and white in rightmost columns), as specified on the axis of the figure. In particular, the leftmost columns of Figs. 15 and 16 display the percent composition of secondary structure of each peptide as a function of r_{ee}. The rightmost columns of Figs. 15 and 16 present the secondary structural make-up of each amino acid residue of the respective peptides as a function of r_{ee}.

The model peptide ALA_{10} displays a clear helix-to-coil transition for the c27 and c36 mechanical unfolding simulations. For the ALA_{10} c27 ASMD simulation, the α-helix is the primary secondary conformation, at least 40 % composition, up to approximately 23 Å of extension. After which the secondary structure transforms primarily into a mix of turn and coil for 5 Å, and then becomes a complete random coil. A similar unfolding pathway is observed for the c36 FF, with the primary difference in pathway being the duration of

the α-helix secondary structure during extension. For the c36 FF ASMD simulation, the α-helix unravels much faster, giving way to the random coil. Both the c27 and c36 ASMD unfolding pathways give rise to a very small degree of 3_{10}-helical characteristics with $i \rightarrow i + 3$ hydrogen bond contacts, predominately occurring centrally within the model helical peptide. The presence of these 3_{10}-helical contacts during the helix-to-coil transition supports a proposed mechanism for helix denaturation which involves passing through a 3_{10}-helix like conformation [153, 154, 151]. The mechanical unfolding of ALA_{10} in the c22 FF however exhibits a distinctively different pathway. The initial helical conformation is not entirely α-helical, as it was in the c27 and c36 ASMD simulations. The initial structure of ALA_{10} in the c22 FF has a significant proportion of π-helical characteristics, which have $i \rightarrow i + 5$ contacts. The π-helical non-native contacts occur toward the C-termini, and then rapidly become frayed to a disordered structure at $\approx r_{ee}$ of 19 Å . Unlike the mechanical unfolding pathway of ALA_{10} in the c27 and c36 FFs, the ASMD simulation in the c22 FF does not exhibit any significant degree of 3_{10}-helical characteristics, but rather the turn structural motif. In the c22 FF, the presence of the π-helix and turn predominate until the peptide is completely unstructured. For the ASMD experiments of our model peptide ALA_{10}, there is a mixed preference for the mode of mechanical unfolding. Meaning for the c27 and c36 FFs simulations there is mechanical unwinding progressing from the C-terminus to the N-terminus. There is a slight difference in the mechanical unfolding pathway of 1PEF in the c27 and in the c36 FF. In the c27 FF the mechanical unfolding occur lastly towards the C-termini in residues 12-18, whereas in the c27 FF the final mechanical unfolding of the α-helix occurs slightly more centrally, affecting residues 8-14. For the c22 ASMD mechanical unfolding, the ALA_{10} appears to occur from both the N- and C-termini simultaneously, leaving the central area of the peptide structured briefly.

Across all FFs, 1PEF maintains primarily α-helical secondary structural components up to an r_{ee} of 52 Å. Because of the strong propensity of 1PEF to retain α-helical contacts, the induced unfolding pathway of 1PEF remains highly structured. This retention of α-helical secondary structure and the similarity of mechanical unfolding pathway across all FFs is illustrated with the positive and consistent slope across both v_s, demonstrates a positive

correlation between the energetics and r_{ee}, which directly correlates to the observed secondary structure along 1PEF's induced unfolding pathway (see Sec. 3.3.2.2). Furthermore, as in the ALA_{10} mechanical unfolding pathway, 1PEF's initial α-helical conformation progressively frays to a more disordered structure, with the loss of internal secondary structure. For the ASMD simulations in the c27 and c36 FFs, the loss of secondary structure is generally proceeding from the N-terminus toward the C-terminus, which is the opposite for what was observed in the ASMD simulations of the model peptide ALA_{10}. Although a variety of mechanisms are involved in the breaking of helical hydrogen bonds, the formation of transient turn structures and structures with mixed α- and turn, or α- and 3_{10} structure is a common motif observed along the 1PEF mechanical unfolding pathway. Primarily for the c22 and c27 forced unfolding simulations, the formation of the 3_{10}-helical contacts occurs more centrally within the 1PEF peptide, starting at an r_{ee} of 42 Å and ending at an r_{ee} of roughly 50 Å. The c36 1PEF mechanical unfolding pathway exhibits similar 3_{10}-helical contacts starting at an r_{ee} of 42 Å and ending at an r_{ee} of roughly 50 Å. But additionally, and uniquely to this FF, 3_{10}-helical contacts are found in the initial starting structure of 1PEF, located near the N-termini. Likewise to ALA_{10}, the c22 initial 1PEF structure is not entirely α-helical, and exhibits π-helical contacts. Albeit 1PEF's π-helical contacts consist of a smaller percent composition of the peptide, and persist for a much shorter duration in the unfolding pathway when compared to ALA_{10}. Furthermore, the mechanical unfolding of 1PEF in the c22 FF proceeds with a loss of secondary structure from the C-terminus towards the N-terminus. This behavior is opposite what is observed during the mechanical unfolding of 1PEF in the c27 and c36 FFs.

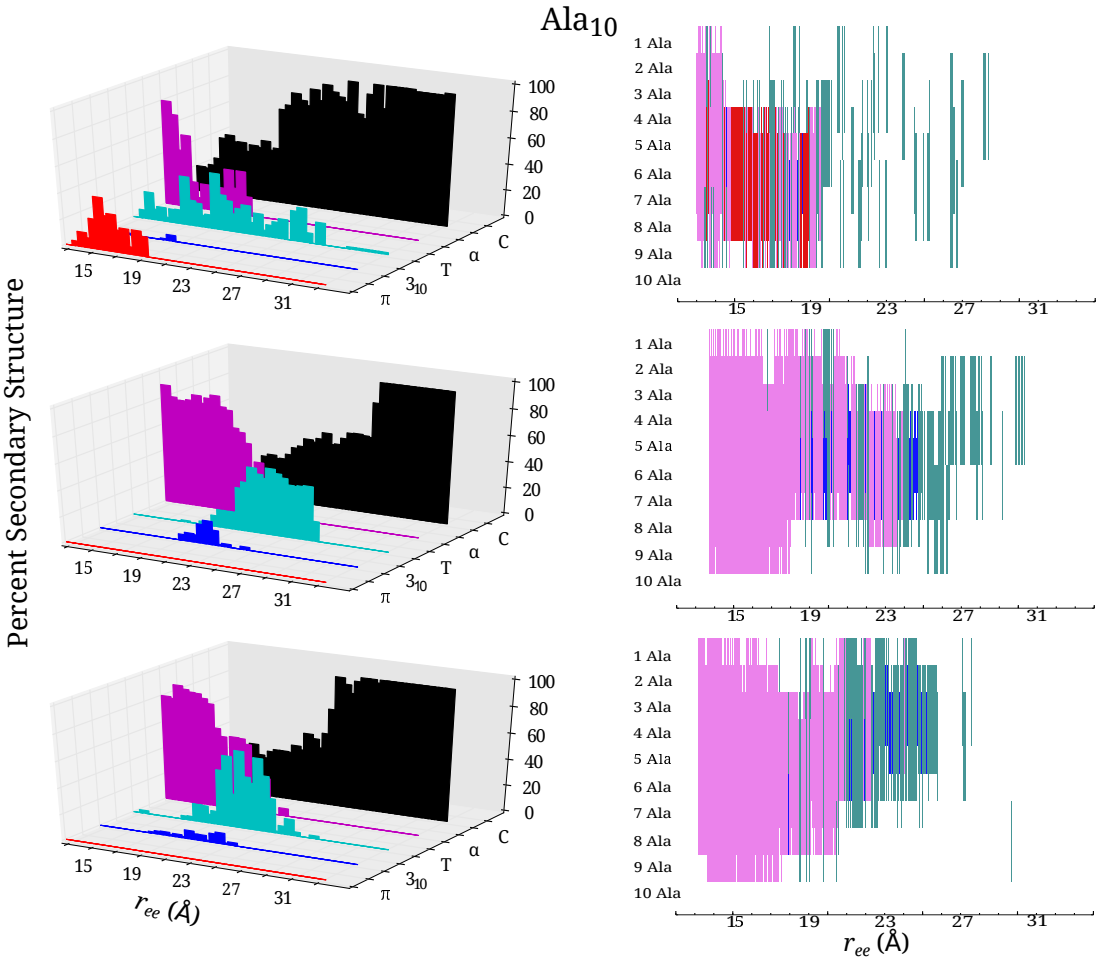

Figure 15: Secondary structure content and propensity was assessed with the STRIDE algorithm in VMD for the model helical peptide ALA_{10} as a function of extension, r_{ee}. The trajectory for the protein in explicit solvent obtained at 1 Å/ns with ASMD that was nearest to the computed Jarzynski average was chosen for analysis in c22 (top panels) in c27 (middle panels), and in c36 (bottom panels). The evolution of the secondary structure character as a function of mechanical unfolding is displayed as: α-helical (purple), π-helical (red), 3_{10}- helical (blue), turn (teal), and random coil (black). The left (right) column of figures displays the percent composition of secondary structure (the secondary structural components of each amino acid residue) as a function of r_{ee}.

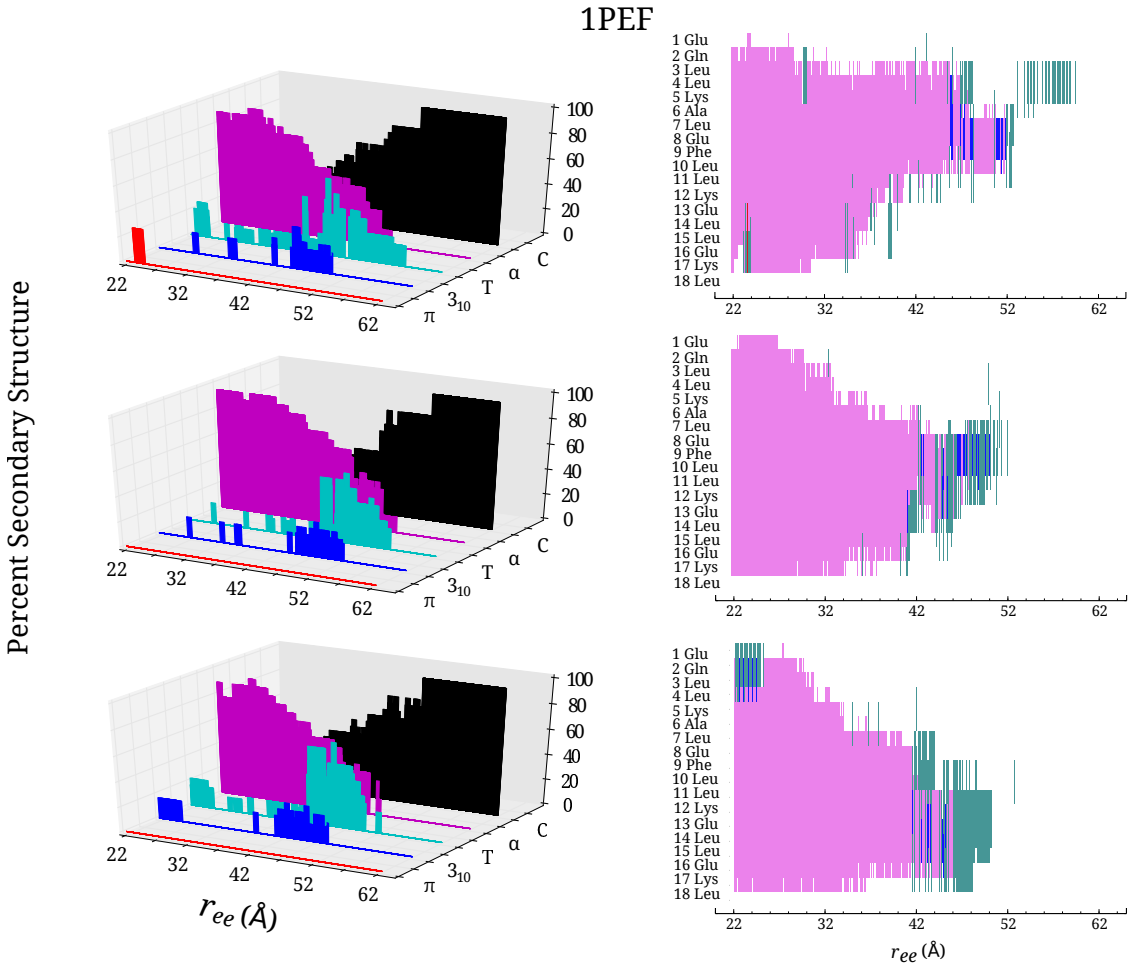

Figure 16: Secondary structure content and propensity for 1PEF is shown as in Fig. 15.

3.3.2.4 Hydrogen Bonding

The converged PMF shown in Fig. 14, and the secondary structural analysis shown in Figures 15 and 16 reveal important structural properties of the helix-to-coil transition of ALA_{10} and 1PEF in water across the family of CHARMM potentials. To more clearly examine the helical preferences and other structural properties of the peptides along the PMFs, a hydrogen bond analysis of the non-equilibrium ASMD ensembles was preformed. The instantaneous number of intrapeptide hydrogen bonds, $\hat{N}_h(S_1, S_2)$, was obtained using the weighting scheme described in Chapter 2. For Fig. 17, $N(=100)$ is the number of trajectories for trajectory i at the extension ζ_t of the peptide backbone.

The hydrogen bonding profiles across each FF seen in Fig. 17, although weighted in

accordance with the work values, still have significant fluctuations. The hydrogen bond profiles for the mechanical unfolding of ALA_{10} are on the left hand side, and the hydrogen bond profiles for the mechanical unfolding of 1PEF are on the right hand side of Fig. 17. The top panels are for the mechanical unfolding using the c22 FF, the central panels are the profiles from the c27 FF, and the last panels are for the ASMD simulations in the c36 FF. The sum of hydrogen bonds for a given structure along the pulling coordinate was partitioned into sums of those hydrogen bonds linking residues separated by a given number of residues along the peptide backbone. In Fig. 17, the $i \rightarrow i + 4$ hydrogen bonds (the blue curves) correspond to those bonds an α-helix. Additionally, the $i \rightarrow i + 3$ hydrogen bond contacts (the red curves) and $i \rightarrow i + 5$ hydrogen bond contacts (the green curves) correspond to those bonds observed in an 3_{10}-helix and π-helix respectively.

For ALA_{10}, the average number of hydrogen bonds seen in the initial compact form of the peptide varies across the FFs. In the c22 ASMD simulations (top panel, left hand side of Fig. 17), the initial compact conformation has a mixture of four $i \rightarrow i + 4$ contacts and one to two $i \rightarrow i + 5$ hydrogen bonds. As was previously discussed and shown in Fig. 15, the π-helical contacts observed in the c22 FF are centrally located within the compact conformation, and are present until an r_{ee} of 19 Å. Additionally, the hydrogen bond analysis of ALA_{10} in the c22 FF shows a reorganization and reformation of α-helical contacts very suddenly during the initial mechanical stretching to an r_{ee} of 15 Å. This reformation of the $i \rightarrow i + 4$ contacts increases the α-helical character of the ALA_{10} peptide, from having initially only four contacts to a maximum of six contacts. After this initial reformation of $i \rightarrow i + 4$ contacts, the α-helical character of the peptide in c22 drops dramatically at an r_{ee} of 19 Å. This dramatic drop in α-helical character corresponds directly to the formation of 3_{10}- and π-helical contacts. After this sudden drop in α-helical character, there is little to no restructuring of the intrapeptide hydrogen bond contacts. The loss of these weak hydrogen bonds, and the lack of secondary structure during the second half of the overall PMF, gives rise to a small free energy increase. For ALA_{10} in c27, (middle left hand side panel) there exists approximately six $i \rightarrow i + 4$ contacts in the starting-compact helical structure. The previously reported α-helical bias of the c27 FF becomes more clear

with the hydrogen bond analysis, in combination with the secondary structural analysis in Sec. 3.3.2.3. These six α-helical contacts are fairly stable and therefore present, on average, to approximately an r_{ee} of 19 Å, after which there is a cross-over to predominately $i \rightarrow i + 3$ character, see Fig. 17. After the model peptide has reached an r_{ee} of 19 Å the secondary structure is dominated by 3_{10} and turn motifs, but some α-helical character remains until an extension of 23 Å. The c27 FF clearly has the most α-helical character observed for the model system across the family of CHARMM potentials. The PMF for the model peptide in the c27 FF (green curve in Fig. 14) has the minimum energy conformation with the lowest energy and demonstrates a free energy landscape with the highest energy plateau. This plateau in the free energy surface represents the extension past 19 Å that has the most α-helical character of any of the FFs, with the greatest energy bias. Lastly, the hydrogen bond profile for ALA_{10} in the c36 FF closely resembles that of the hydrogen bond profile from the c27 FF, with one major exception; the total number of α-helical contacts along the pulling coordinate are fewer. ALA_{10} in the c36 FF has five $i \rightarrow i + 4$ contacts in the compact structure, and this number declines rapidly upon the peptide unravelling. On the other hand, ALA_{10} in the c36 FF become characterized by equal number of α-helical and 3_{10}-helical contacts at a cross-over point on the pulling coordinate with an r_{ee} of about 19 Å. This cross-over point matches the cross-over of secondary structure observed in the c27 simulations. At this cross-over point there are an equal number of α-helical and 3_{10}-helical hydrogen bonds present in the ensemble. All α-helical contacts, across all FFs for the mechanical unfolding of ALA_{10} are ruptured at around 23 Å. The ASMD hydrogen bond profiles of ALA_{10} are further substantiated with the secondary structural analysis in Sec. 3.3.2.3. There are no π-helical contacts observed during the peptide extension in the c27 or c36 FF as there was in the c22 FF. The α-helical contacts are mostly replaced by 3_{10} contacts within the first 10 Å of peptide backbone extension in c27 and in c36 at which the cross-over point is reached. This would explain how the energy in the PMF, see Fig. 14, remains nearly constant and flat up to about 23 Å r_{ee}, where the peptide becomes completely unstructured.

According to the XRD structural information of 1PEF, twelve of the fourteen main chain

intra-molecular hydrogen bonds are α-helical in character and the last two main chain intra-molecular hydrogen bonds have 3_{10}-character. The initial conformation of 1PEF for each ASMD simulation is a compact helix with unspecified helical contacts with an r_{ee} of 22 Å. Similarly to ALA_{10}, 1PEF displays π-helical contacts in the c22 ASMD pulling experiments, resulting with a shorter and fatter helix. These π-helical contacts shift the minimum energy r_{ee} to the right, making the energy minimum occur at 24.06 Å (with a well depth of 3.2 kcal/mol, refer to Fig. 14). Furthermore, just like ALA_{10} in the c22 FF, there are very little to no 3_{10}-helical contacts in the compressed or minimum energy structure. The mechanical unfolding pathway experienced by 1PEF in the c27 and c36 FF are very similar to one another, just as the two pathways were similar to one another in the two FFs for ALA_{10}. There is a clear shift of the hydrogen bonds from α-helical to 3_{10}-helical for the c27 and c36 FF ASMD simulations of 1PEF, occurring near 45 Å of extension. Additionally, there is a slight reorganization of hydrogen bonds in the c27 and c36 FF for the mechanical unfolding of 1PEF. The restructuring of these hydrogen bond contacts increase the number of α-helical contacts within the peptide at the minimum energy structure observed along the PMF. The minimum energy structure of 1PEF in the c27 FF occurs with an r_{ee} of 25.93 Å at a well depth of 4.1 kcal/mol. In c36, the minimum energy conformation occurs with an r_{ee} of 25.80 Å at a well depth of 3.5 kcal/mol. The c27 ASMD simulation shows the over stabilization of the α-helical peptide with a constant 12 - 14 α-helical contacts formed, and no 3_{10} contacts. Likewise in the c36 FF, during the mechanical unfolding of 1PEF, there is a reorganization of the α-helical hydrogen bonds increasing the overall number within 1PEF. But the duration of the maximum number of hydrogen bonds is shorter than in the c27 FF. As further reflected in Fig. 16, the secondary structure of 1PEF remains α-helical until $\approx$ 42 - 44 Å r_{ee} for both the c27 and c36 FF. However, the unravelling in the c36 FF occurs more towards the C-termini in residues 12-18, whereas in the c27 FF the final unravelling of the α-helix occurs slightly more centrally, residues 8-14. For 1PEF, the α-helical contacts are replaced by 3_{10} contacts within the last 10 Å of peptide backbone extension in c27 and in c36 at which the cross-over point is reached. However, unlike ALA_{10}, the α-helical bonds are not completely replaced by 3_{10}-helical contacts in the c27 or c36 FF.

This would explain how the energy in the PMF, see Fig. 14, remains constantly increasing and flat up to about 52 Å r_{ee}, where the peptide becomes completely unstructured. For our model system, all secondary structure (excluding turn motifs) is lost after approximately 15 Å of steering the peptide backbone. Whereas for 1PEF, all secondary structure, again excluding turn motifs, is lost after approximately 30 Å of steering the peptide backbone. This provides an explanation for the cross-over point occurring roughly at about twice as far along the pulling coordinate; 19 Å for ALA_{10} and approximately 45 Å for 1PEF. The c36 results are about the same as the c27 results; the free energy results in the c27 FF are higher than the results from the c36 FF. The similarities of the mechanical unfolding pathways between our model system ALA_{10} and the biological relevant system 1PEF demonstrate how the corrections from the CMAP reparameterization take affect.

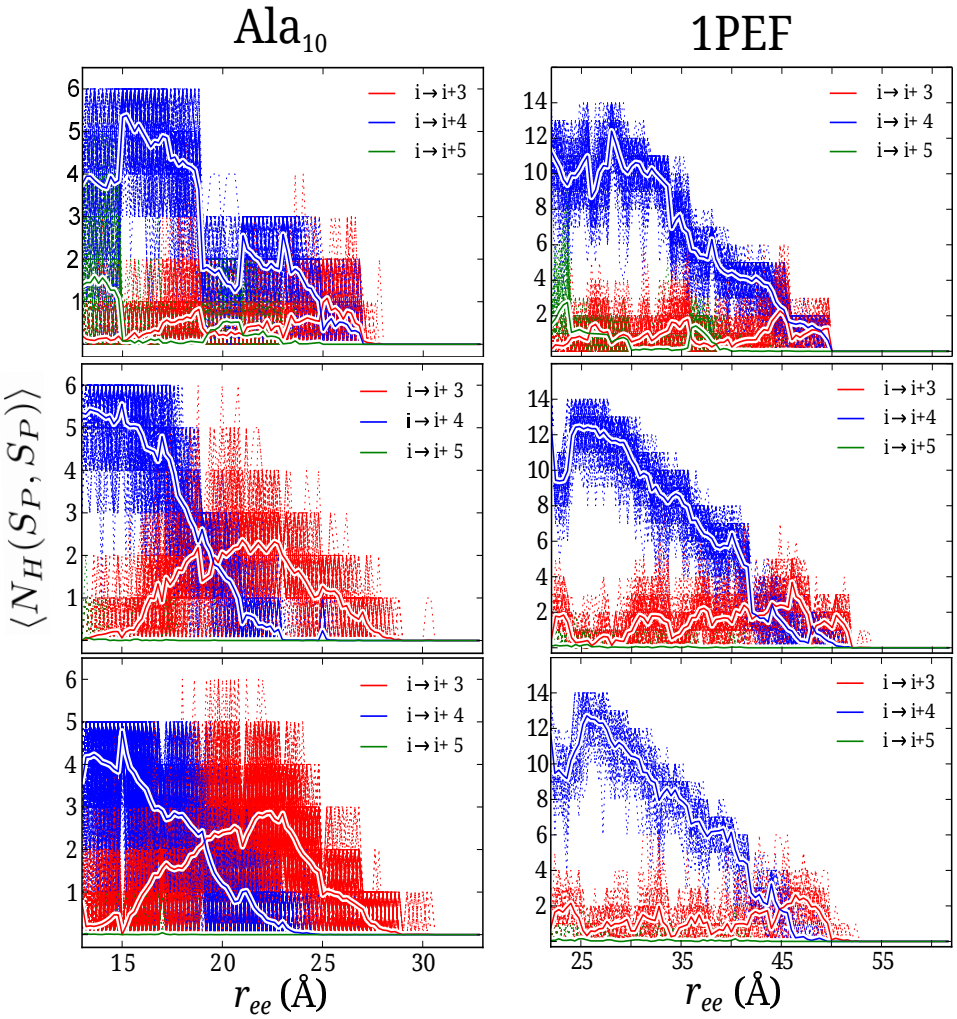

Figure 17: The average number of α-, 3_{10}-, and π-helical contacts as a function of the end-to-end distance r_{ee} are shown for the mechanical unfolding of ALA_{10} (left) and 1PEF (right) in explicit solvent using the c22 (top), c27 (middle), and c36 (bottom) FFs as obtained by ASMD with $v_s = 1$ Å/ns with 100 tps. Red curves represents The $i \to i+3$ (3_{10}-helix), $i \to i+4$ (α-helix) and $i \to i+5$ (π-helix) contacts are shown in red, blue and green, respectively.

3.3.3 Comparing the performance of CHARMM36 and Amber16 using the helical motif

3.3.3.1 Energetics

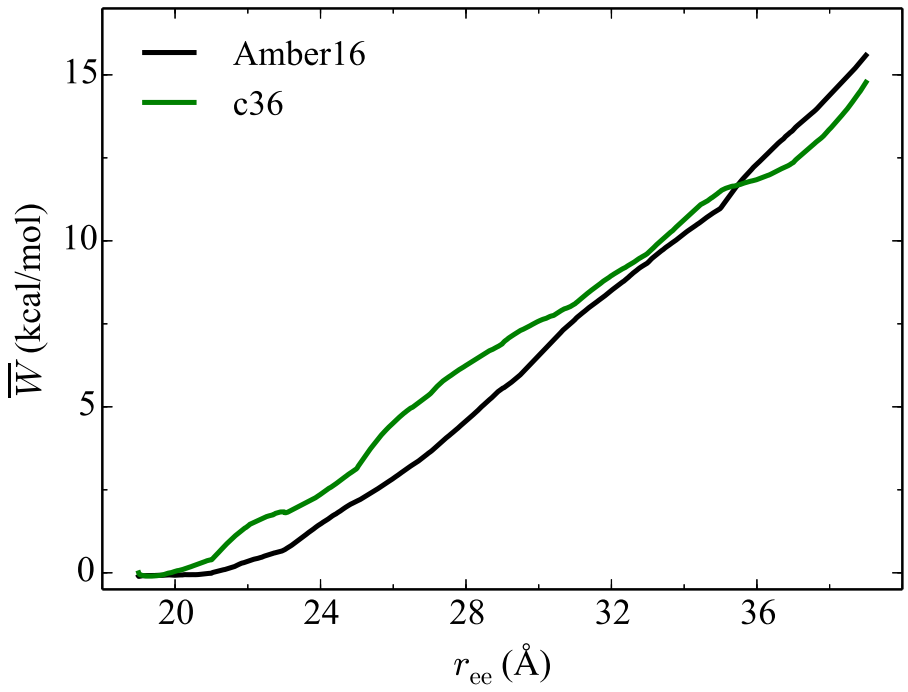

Figure 18: Energetic comparison of $gp41_{659-671}$ in both c36 and Amber16. The PMFs are obtained using ASMD with 100 tps at a pulling velocity of 1 Å/ns.

The PMFs of two peptides are used as comparative systems to test the performance of the c36 and Amber16 FFs. The energetics of the forced unfolding of $gp41_{659-671}$ and 1PEF are shown in Figs. 18 and 19, respectively. The energetics were obtained using ASMD with 100 trajectories per stage using 1 Å/ns. The PMFs have several similarities and few differences between the two FFs. In the $gp41_{659-671}$ results, shown in Fig. 18, both PMFs have the same minimum at a r_{ee} of 19.7 Å. Slight differences emerge in the PMFs in the region between 22 and 32 Å. In this region the c36 PMF (black curve) maintains a slightly higher energy, though it is with thermal fluctuation of the Amber16 result (green curve). At a r_{ee} of 36 Å, the Amber16 result become slightly more stabilizing yet still within thermal

fluctuation of the c36 result. The $gp41_{659-671}$ PMFs are both approximately equivalent in total unfolding energy.

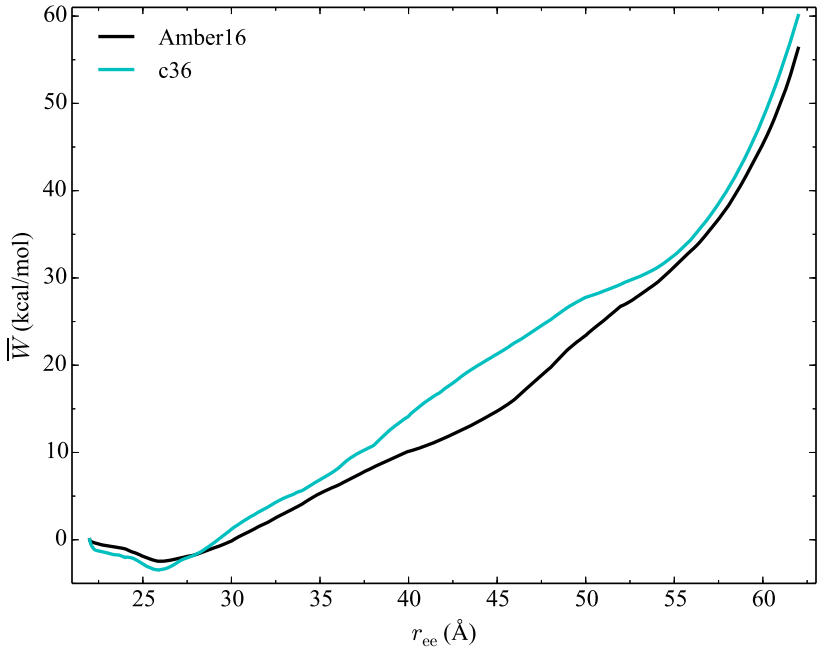

Figure 19: Energetic comparison of 1PEF in both c36 and Amber16. The PMFs are obtained using ASMD with 100 tps at a pulling velocity of 1 Å/ns.

Similar observations can be made about the 1PEF PMFs shown in Fig. 19. The PMFs have similar minima regions between 22 and 26 Å. Again, the c36 PMF (cyan curve) shows a slight increase in stabilization in the region between 27 and 55 Å. Within that region, the maximum difference between the PMFs is $\approx$ 6 kcal/mol at a r_{ee} of 47 Å. This difference is significant as it is larger than the thermal fluctuation. Overall, the magnitude and curvature of the PMFs are consistent with one another. These results suggest that the use of different potentials does not significantly effect the calculation of the forced unfolding energetics of helical peptides.

3.3.3.2 Residue pair interaction energies

1PEF 1PEF has an amino acid sequence that contains charged residues. Therefore, in addition to hydrogen bonding, long range electrostatic interactions can occur and increase the stabilization of the α-helix. These salt bridges can be formed by either ionizable residue side-chains, or between charged groups, or polar groups in amino acid side chains. The sequence and possible ionic pairs for 1PEF are shown in Fig. 20. (where ACE represents the acetylated N-terminus and CT2 the amidated C-terminus), 1PEF was a *de novo* designed peptide, and has a high apolar-polar amino acid ratio (10:8) [93]. The amphiphilic nature of the helix surface results in one surface being hydrophobic and the other being hydrophilic. The hydrophilic surface of 1PEF is clearly illustrated in Fig. 20, in which the charged residues are highlighted: blue for cationic residues and red for anionic residues. 1PEF was purposefully designed in hopes to enhance the potential for interhelical interactions within the self-assembly of helical bundles or aggregates of 1PEF. Specifically, within the design of the α-helix, 1PEF has 8 apolar amino acids in two neighboring rows of four on one surface of the peptide: Leu3-Leu4, Ala6-Leu7, Leu10-Leu11, and Leu14-Leu15. This row of hydrophobic residues on one face of the peptide assisted in creating a peptide that has a high degree of helicity and self-associates into hexamers. Therefore, the oligomerization is in part due to hydrophobic packing of "ridges-into-grooves" [93]. More interestingly, 1PEF can oligomerize in both hydrophobic and hydrophilic environments, which raises the question as to how this more balanced amphililicity character affects the self-assembly. The polar interface of 1PEF (Fig. 20) was found to be involved in a network of complex water-mediated hydrogen bonds and salt bridges. The polar residues must play a key role in the stability of the secondary structure of 1PEF through salt bridges. In particular, three salt bridges are observed experimentally: Lys5-Glu8, Lys12-Glu13, and Lys12-Glu16 [93]. For these ionic pairs, the basic residue is closer to the amino terminus with respect to the acidic amino acid. Additionally, it was predicted by Marqusee and Baldwin [155] that for a single, non-interacting α-helical peptide, salt bridges would form with the acidic residue toward the amino terminus with respect to the basic amino acid. It is predicted that this pattern of ionic pairing is more stable, because the dipoles of the salt bridges have a more favorable interaction with the helix dipole; the positive pole of the helix is near the N-terminus and

the negative pole is near the C-termini [155]. In 1PEF such salt bridges would form between the following pairs: Glu1-Lys5, Glu8-Lys12, and Glu13-Lys17. The arrows in Fig. 20 are a cartoon representation of the possible salt bridge interactions.

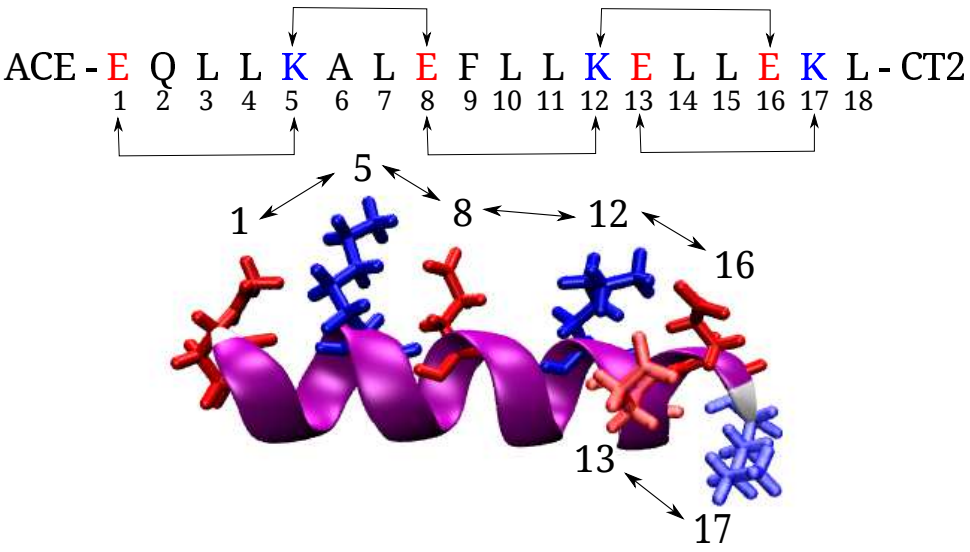

Figure 20: The primary amino acid sequence and the cartoon secondary structural representation (a purple ribbon) of 1PEF is shown. The positive amino acid lysine (K) is colored blue, and the negative amino acid glutamic acid (E) is colored red. The arrows illustrate the ion pairs investigated; i.e., Lys5-Glu8 which is an $i \rightarrow i + 3$ contact. The $i \rightarrow i + 1$ ion pair arrows are omitted for clarity (Lys12-Glu13 and Glu16-Lys17).

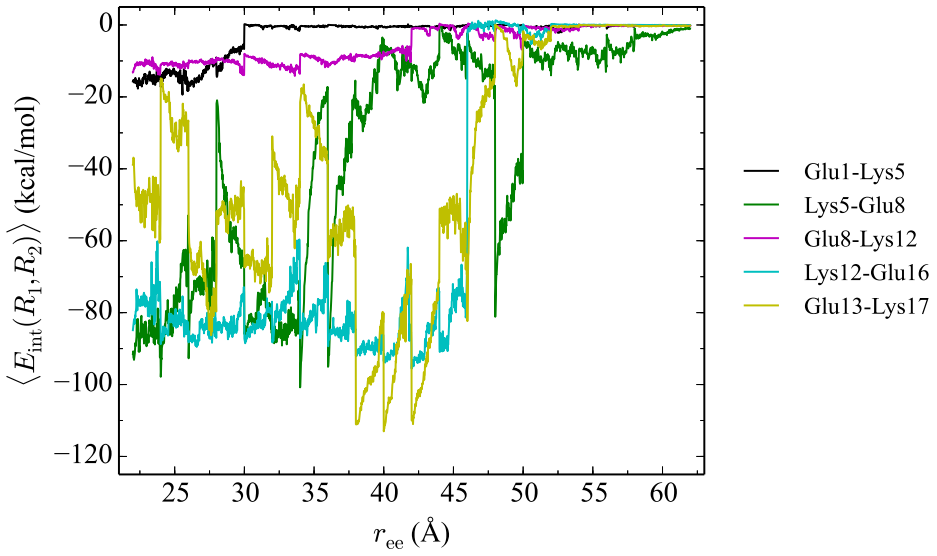

Figure 21: The residue pair interaction energies for 1PEF obtained using c36. The curves are obtained from ASMD simulations using 100 tps at 1 Å/ns using explicit water solvent. The residue pairs of interest are Glu1-Lys5 (black curve), Lys5-Glu8 (green curve), Glu8-Lys12 (magenta curve), Lys12-Glu16 (cyan curve) and Glu13-Lys17 (yellow curve) as denoted in the legend.

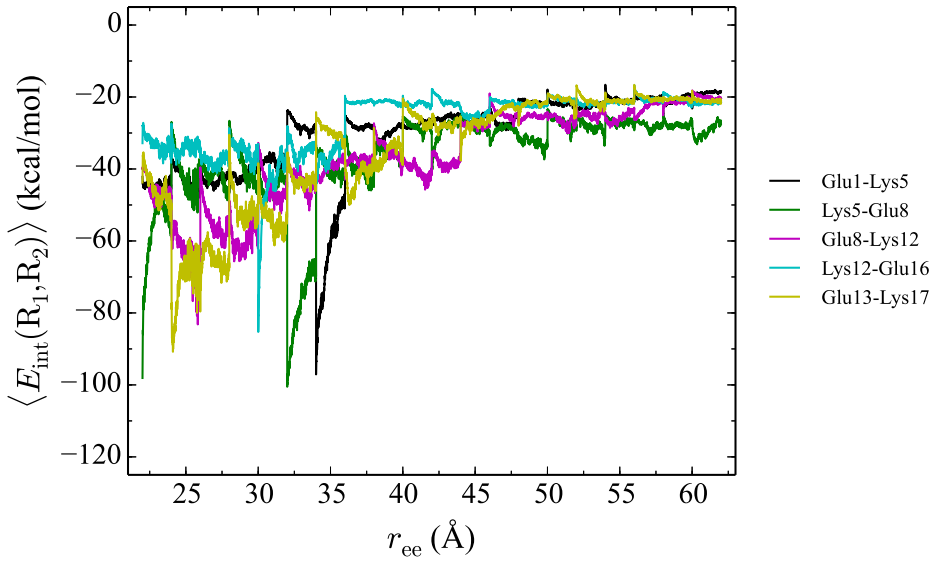

Figure 22: The residue pair interaction energies for 1PEF obtained using Amber16. The curves are obtained from ASMD simulations using 100 tps at 1 Å/ns using explicit water solvent. The residue pairs of interest are Glu1-Lys5 (black curve), Lys5-Glu8 (green curve), Glu8-Lys12 (magenta curve), Lys12-Glu16 (cyan curve) and Glu13-Lys17 (yellow curve) as denoted in the legend.

The hydrophilic face of 1PEF is comprised primarily of EK or KE type salt bridges. Both the location and the spacing of charged substituents along a helical peptides backbone will greatly influence the peptide's helical content and the overall stability of the peptide. Appropriately spaced salt bridges can increase the stability of an α-helix by anchoring the secondary conformation into place across individual helical turns [155, 94]. Therefore, we have evaluated the effective interaction energies between different pairs of residues along $r_{\rm ee}$ as shown in Figs. 20 and 22. The interaction energies are weighted averages obtained using the methods described in Chapter 2. For 1PEF, several residue pairs were identified and monitored over the course of the unfolding: four $i \to i+4$ contacts (Glu1-Lys5, Glu8-Lys12, Lys12-Glu16, and Glu13-Lys17) and one $i \to i+3$ contact (Lys5-Glu8).

In the comparison of the two FFs, the interaction energies of the salt bridges in Amber16 are continually stabilized throughout the entire pull. Moving sequentially along the

backbone of 1PEF the first ionic pair is Glu1-Lys5, an $i \rightarrow i + 4$ interaction (black curves in Figs. 21 and 22). This interaction is located near the positive pole of the helix, which was predicted to increase the stability of the helix [155]. However, this particular salt bridge plays a insignificant role in stabilizing the peptide along the mechanical pathway of unfolding in c36. However, in Amber16 this interaction begins at ≈ -40 kcal/mol and is further stabilizing at a r_{ee} of 35 Å but by the end of the pull is decreased to -20 kcal/mol. The next pair of interest is the $i \rightarrow i + 3$ interaction Lys5-Glu8 (green curves in Figs. 21 and 22). The Lys5-Glu8 ionic pair is experimentally observed to be a stabilizing interaction [93]. Likewise through the ASMD simulations, the Lys5-Glu8 pair shows a stronger, favorable interaction particularly in c36. 1PEF mechanically unfolds from the N-terminus to the C-terminus, and as a result, the Lys5-Glu8 salt bridge contact appears through an r_{ee} of ≈ 35 Å. A majority of the 3_{10} helical contacts exist after an extension of 35 Å r_{ee} and are mostly found in the central part of the peptide, and not in the N-terminus. The Glu8 residue is able to rotate from this $i \rightarrow i + 3$ contact, to an $i \rightarrow i + 4$ contact (Glu8-Lys12, magenta curve in Fig. 21). The ionic pair Glu8-Lys12 was predicted to have stabilizing effects for the single, non interacting 1PEF peptide [155]. In c36, Glu8-Lys12 appears to have little to no energetic contribution to the unfolding pathway, and thus the Glu8 residue is primarily involved in the $i \rightarrow i + 3$ Lys5-Glu8 contact in c36. This is not the case in Amber16. The Glu8-Lys12 interaction is stabilized for the entirety of the pull. The Lys5-Glu8 begins as stabilizing but drops off within the first stage of the pull as shown in green in Fig. 22. The Glu8-Lys12 interaction then has the ability to become the most stabilizing of the two competing interactions.

For the last ionic pairs observed in 1PEF, the Lys12-Glu16 pair has been investigated experimentally [93]. The Lys12-Glu16 pair is an $i \rightarrow i + 4$ contact (cyan curve in Figs. 21 and 22). For this experimentally observed salt bridge pair, the Lys12 residue can rotate similar to the Glu8 residue in the previous pairs. The Lys12-Glu16 pair has favorable interaction in the c36 FF. This pair shows considerable stability and longevity between an r_{ee} of ≈ 22-45 Å. Again, this interaction is captured differently by the Amber16 FF. In Fig. 22, the Lys12-Glu16 interaction is initially the least stabilized of all of the interactions

at -30 kcal/mol. By the end of the reaction coordinate, the interaction has been destabilized to -20 kcal/mol. The last salt bridge pair of interest occurs closest to the C-termini, Glu13-Lys17 (yellow curve in Figs. 21 and 22.) In similar fashion to the last ionic pair discussed (Lys12-Glu16), this pair shows favorable interaction energies in the c36 FF. The $i \rightarrow i + 4$ pair, Glu13-Lys17, is the only predicted ionic pair [155] to show any substantial stabilizing effects. In c36 the Glu13-Lys17 pair maintains stability but has significant fluctuations. In c36, they have highly favorable energetic fluctuations until $\approx r_{ee} = 45$ Å. The mechanical unfolding of 1PEF in c36 exhibits the most α-, 3_{10}-, and turn characteristics out of the three FFs closet to the C-termini.

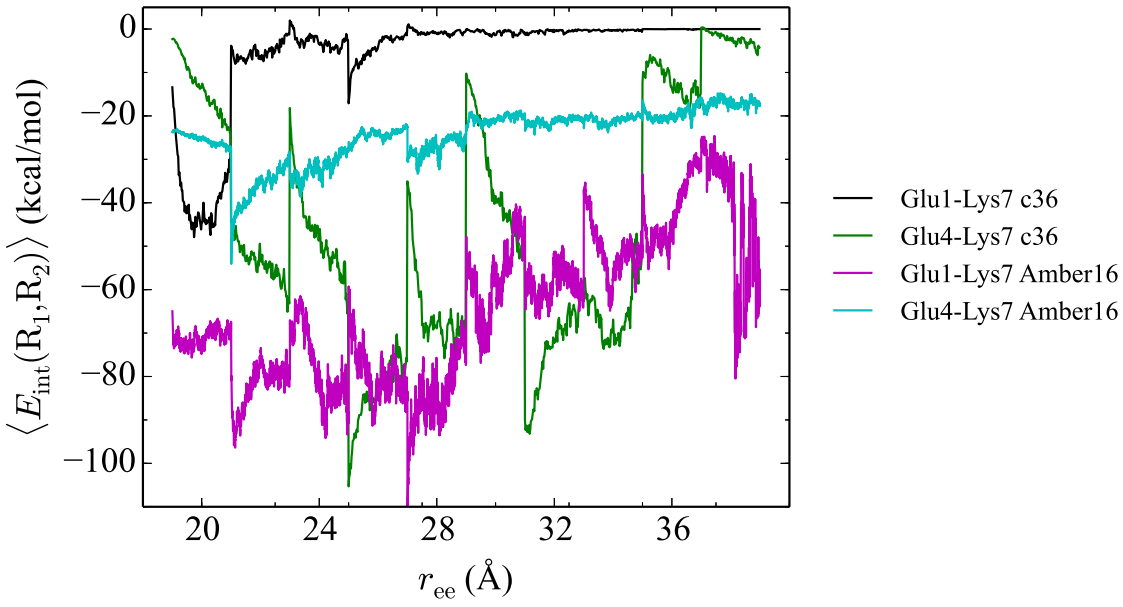

Figure 23: Comparison of the residue pair interaction energies of two salt bridge interactions within $gp41_{659-671}$. The curves are obtained from ASMD simulations using 100 tps at 1 Å/ns using explicit water solvent. The residue pairs of interest are Glu1-Lys7 in c36 (black curve), Glu4-Lys7 in c36 (green curve), Glu1-Lys7 in Amber16 (magenta curve) and Glu4-Lys7 in Amber16 (cyan curve) as denoted in the legend.

gp41$_{659-671}$ Similar to 1PEF, $gp41_{659-671}$ contains residues that have the dynamic possibility of forming salt bridges. The residue pair interaction energies for two salt bridge pairs

within $gp41_{659-671}$ are shown in Fig. 23. The two salt bridges of interest are Glu1-Lys7 and Glu4-Lys7. These pairs of salt bridges share the Glu1 residue and are therefore competing. This makes them excellent choices with which to analyze and compare the performance of the two FFs. In c36, the Glu1-Lys7 (black curve) is only stabilizing in the beginning of the unfolding from 19 to 21 Å. The interaction is effectively nonexistent for the remainder of the pull as the residues are no longer within the range of interaction. However, this is not the trend observed in Amber16. The Glu1-Lys7 interaction in Amber (magenta curve) shows substantially more stabilization over the entire course of the unfolding reaction coordinate. The interaction begins as the most stabilizing of the two pairs investigated. The interaction is damped over the course of the reaction coordinate, but only by $\approx$ 20 kcal/mol and by the end of the unfolding, is still maintaining significant interaction. The reverse trend is observed for the Glu4-Lys7 interaction between the FFs. In c36, this interaction (green curve) becomes very stabilized over the pull. Initially, the interaction is not stabilizing; however, it quickly become the most stabilizing interaction over the first 4 Å of the pull. The interaction undergoes significant fluctuation over the entire unfolding, but maintains contact until the last 4 Å at which the interaction is less stabilizing than the Amber16result. This is in contrast to the interaction in Amber16 (cyan curve) which only shows slight stabilization through the unfolding. The interaction begins at $\approx$ -20 kcal/mol and remains consistent at this energy for the duration of the simulations.

3.3.3.3 Hydrogen bonds

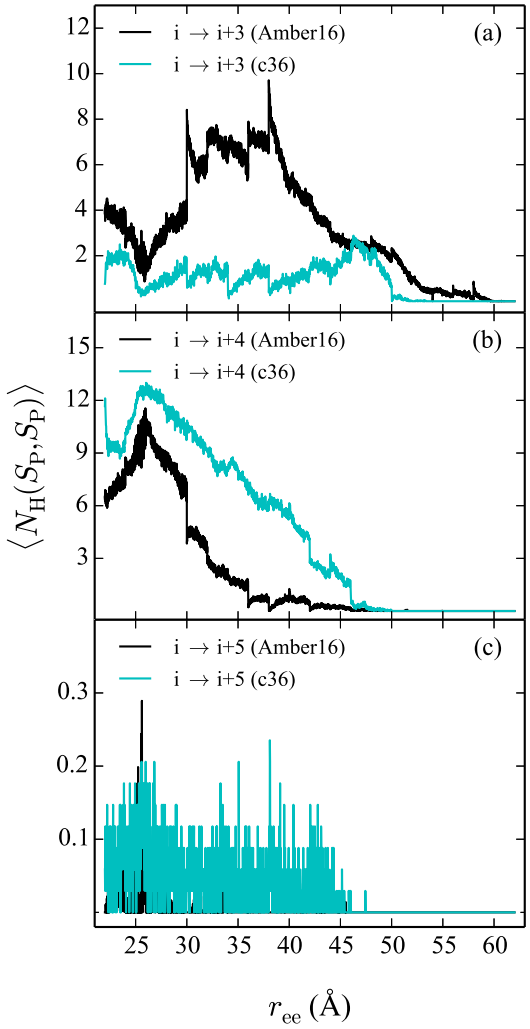

Figure 24: Comparison of the intrapeptide hydrogen bonds formed within 1PEF. The top panel corresponds to the $i \rightarrow i+3$ bonds, middle panel to the $i \rightarrow i+4$ bonds and the bottom panel to the $i \rightarrow i+5$ bonds. All of the hydrogen bond results are the weighted averages of 100 tps at 1 Å/ns. The c36 results are shown in cyan and the Amber16 are shown in black in each panel.

The intrapeptide hydrogen bonds formed and broken within 1PEF during the mechanical unfolding are shown in Fig. 24. In Fig. 24a, the $i \rightarrow i+3$ bonds are shown. These bonds correspond to a 3_{10}-helix. In the c36 result (cyan curve), there is very little 3_{10}-helix formation. This is in contrast to the Amber16 result (black curve) which shows an initial decrease in the number of contacts but in the region from 25 to 35 Åshows a significant increase from 2 to 8 contacts. The number of contacts begins decreasing to 0 at a r_{ee} of 40 Å. The FFs are considerably more consistent in Fig. 24b which depicts the $i \rightarrow i+4$ α-helical bonds formed and broken within 1PEF. Both the c36 and Amber16 patterns show an initial increase in the number of bonds formed until a r_{ee} of $\approx$ 26 Å. At that point, the number of contacts decreases from $\approx$ 11 to 0. However, in general, the c36 result does capture more $i \rightarrow i+4$ bonds than the Amber16 result. Finally, in Fig. 24c, the FFs are consistent in the number of $i \rightarrow i+5$ (π-helical) bonds formed and broken. Both FFs effectively capture no π-helical contacts over the unfolding reaction coordinate.

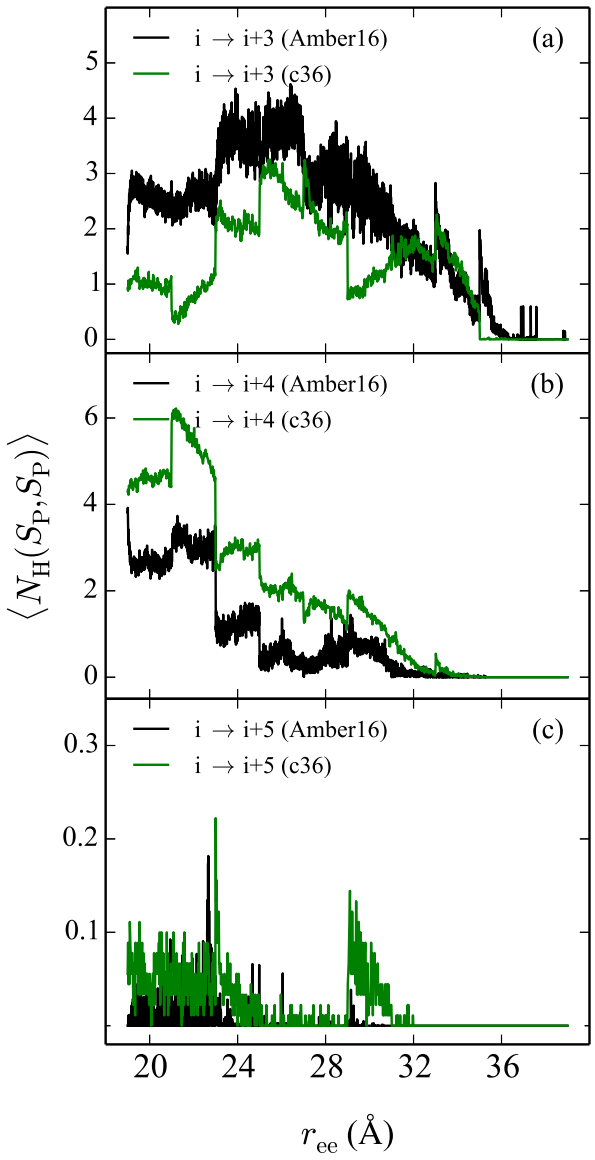

Figure 25: Comparison of the intrapeptide hydrogen bonds formed within $gp41_{659-671}$. The top panel corresponds to the $i \rightarrow i+3$ bonds, middle panel to the $i \rightarrow i+4$ bonds and the bottom panel to the $i \rightarrow i+5$ bonds. All of the hydrogen bond results are the weighted averages of 100 tps at 1 Å/ns obtained using explicit water solvent. The c36 results are shown in green and the Amber16 are shown in black in each panel.

Variation in the type and number of intrapeptide hydrogen bonds captured during the forced unfolding of $gp41_{659-671}$ can be seen in Fig. 25. The $i \rightarrow i+3$ hydrogen bonds are shown in Fig. 25a. The overall trend of an initially lower number of 3_{10} contacts with a steady increase between a r_{ee} of 24 to 38 Å and then an decreasing region between 24 to 36 Å is preserved between the Amber16 and c36 FFs. However, there are slight differences between the average number of hydrogen bonds captured by the individual FFs. The Amber16 result (black curve) captures more $i \rightarrow i+3$ contacts than the c36 result (green curve). Initially, the Amber16 FF begins with 3 contacts and eventually maximizes with 4 contacts before decreasing to 0 contacts by the end of the reaction coordinate at 36 Å. In contrast, the c36 begins with identifying only 1 contact and increasing to 3 contacts before a steady decline to 0. In Fig. 25b, the trend of Amber16 capturing more hydrogen bonds than c36 is reversed. These are the $i \rightarrow i+4$ contacts which correspond to an α-helix. In the c36 result, the hydrogen bond count initially begins at 4 then immediately increases to 6 at a r_{ee} of 22 Å. From that point on, the number of intrapeptide α-helix bonds decreases to 0 by a r_{ee} of 32 ÅThis is consistent with the overall trend captured by the Amber16 FF; however, the number of bonds captured are fewer than in c36. It should be noted that this is consistent with the intrapeptide hydrogen bonds of 1PEF shown in Fig. 24a and Fig. 24b. Lastly, in Fig. 25c, the $i \rightarrow i+5$ hydrogen bonds are shown. Similar to the results with 1PEF, neither Amber16 nor c36 capture a significant number of these contacts, which would indicate formation of a π-helix. From this analysis it is clear that the peptide, regardless of FF used for the dynamics, presents as a mixed helix containing both 3_{10}- and α-helical contacts.

3.3.4 Linearity of elongating alanine-rich peptides

3.3.4.1 Energetics

Preliminary results of the unfolding energetics of a series of polyA peptides have been calculated and compared. The PMFs of all of polyAs of length 6, 10, 14, 22, 30, 38 and 50. The results can be seen in Fig. 26. All of the PMFs of the peptides, with the exception of ALA_6 (red curve), exhibit a trend of increasing linearity. The PMF of ALA_6 does not

follow this trend as the mechanical unfolding results in a PMF that is unexpectedly high as the final value of the PMF is $\approx$ 33 kcal/mol. ALA_6 also contains a second minima located at a r_{ee} of 6 Å. One reason for these differences is that the peptide is not helical, as the other peptides in the series are; it contains significant turn character within the structure. This is corroborated by the results of the hydrogen bonding profile of ALA_6 presented in Sec. 3.3.4.2. Turn character is a precursor to β-hairpin formation. This is consistent with our other studies of the mechanical unfolding of β-hairpin peptides presented in detail in Chapter 4. Our studies suggest that the linear unfolding of β-hairpin peptides results in a PMF that is much greater than the unfolding energetics of helical peptides.

The other polyAs in the series have energetics of unfolding that increase with increasing peptide length as expected. The next peptide in the series, ALA_{10} (blue curve), has a final value of the unfolding PMF of 17 kcal/mol. This is consistent with our previous studies using ALA_{10}, shown in Figs. 6 and 14, despite the difference in the endcaps of the peptide. ALA_{14} (green curve) has a final PMF value of 30 kcal/mol. The increase in length of 4 residues between ALA_{10} and ALA_{14} corresponds to an increase in helical length by one "turn". Therefore, the energetic increase of one turn is approximately equivalent to a difference of 13 kcal/mol. The next peptide in the series, ALA_{22} (black curve) is larger than its predecessor by 2 turns. It's final value of the PMF is $\approx$ 52 kcal/mol. This trend also holds for the increase in PMF between ALA_{22} and ALA_{30} (magenta), which is also an increase in length by 2 turns. The next peptide in the series, ALA_{38} (yellow curve), does not follow this increase rule. The difference between the final PMF values of ALA_{30} and ALA_{38} are negligible. However, the trend emerges again in the difference between ALA_{38} and ALA_{50} (cyan curve), which corresponds to an increase of 3 turns. The difference is $\approx$ 40 kcal/mol.

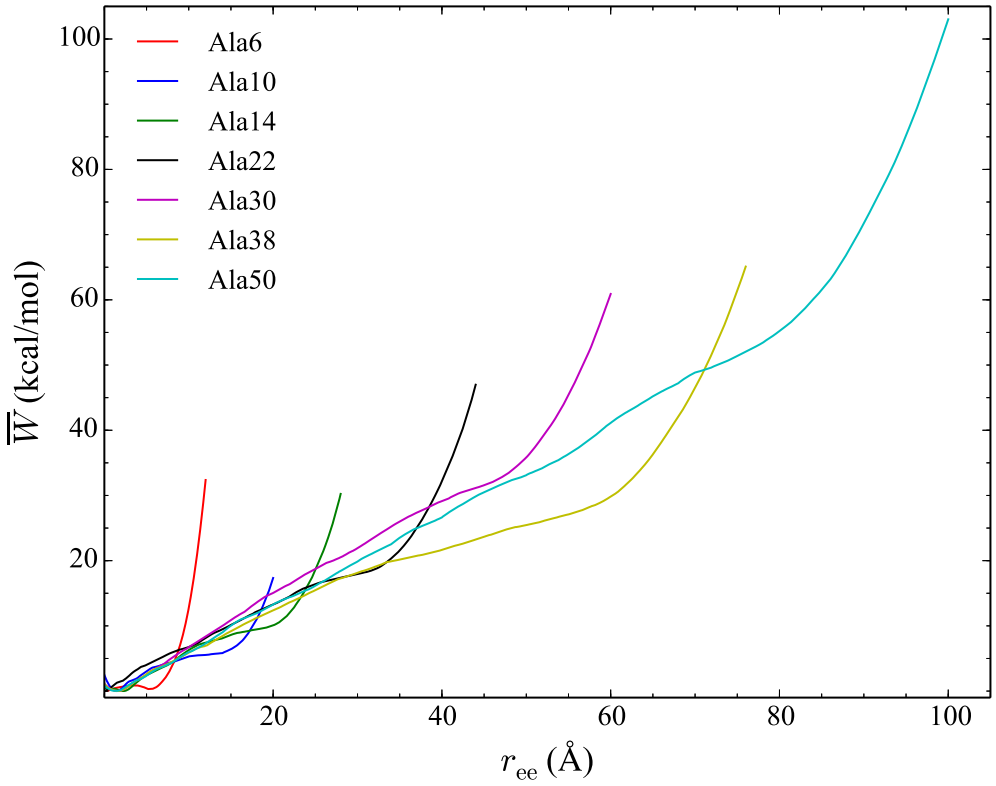

Figure 26: Comparison of the energetics of several polyalanine peptides. The PMFs have been obtained using explicit water solvent at 100 tps at 1 Å/ns in the c36 FF.

3.3.4.2 Hydrogen bond profiles

The hydrogen bonding patterns of seven alanine-rich peptides are shown in Fig. 27. All of the curves are the weighted averages of 100 trajectories per stage using the weighted averaging scheme described in Chapter 2. In this figure, the curves have not been shifted by the initial value of r_{ee} (as the PMFs in Fig. 26) for clarity. In Fig. 27a, the $i \rightarrow i+3$ hydrogen bonds which correspond to formation of a 3_{10}-helix, are are shown for each peptide. No peptide shows substantial formation of 3_{10}-helical contacts. However, an interesting trend does arise: the longevity of the $i \rightarrow i+3$ contacts increases with increasing helix length. For example, ALA_6 (red curve) initially contains 1 $i \rightarrow i+3$ hydrogen bond, but the contact is immediately broken during the first stage of the forced unfolding. ALA_{10} (blue curve)

shows slight formation of 1.5 $i \rightarrow i+3$ hydrogen bonds, but the bonds are broken within the first couple of stages. The ALA_{14} (green curve) forms 1 3_{10}-helical contact which is only transiently stable. The ALA_{22} (black curve) forms and breaks a 3_{10}-helical contact during the unfolding. The single 3_{10}-helical contact formed within ALA_{30} (magenta curve), ALA_{38} (yellow curve) and ALA_{50} (cyan curve) is maintained for a significant portion of the unfolding pathway. In the ALA_{50} curve, the contact is maintained from a r_{ee} of $\approx$ 72 to 148 Å, constituting a majority of the pull. In Fig. 27b, the $i \rightarrow i+4$ hydrogen bonds which correspond to formation of a α-helix, are are shown for each peptide. All of the peptides, with the exception of ALA_6, are predominantly α-helices and the number of α-helical contacts increases linearly with the increasing length of the peptides. ALA_{10} is the smallest of the peptides that contain $i \rightarrow i+4$ hydrogen bonds as it initially begins with ≈ 5 hydrogen bonds. ALA_{50} is the largest peptide; and therefore, contains the initially the highest number of $i \rightarrow i+4$ bonds. In Fig. 27c, the $i \rightarrow i+5$ hydrogen bonds which correspond to formation of a π-helix, are are shown for each peptide. As expected there is virtually no π-helical formation within any of the peptide over the course of the reaction coordinate. These findings are also consistent with our previous ASMD benchmarks using c36 with other helical peptides as described in Sec. 3.3.2.

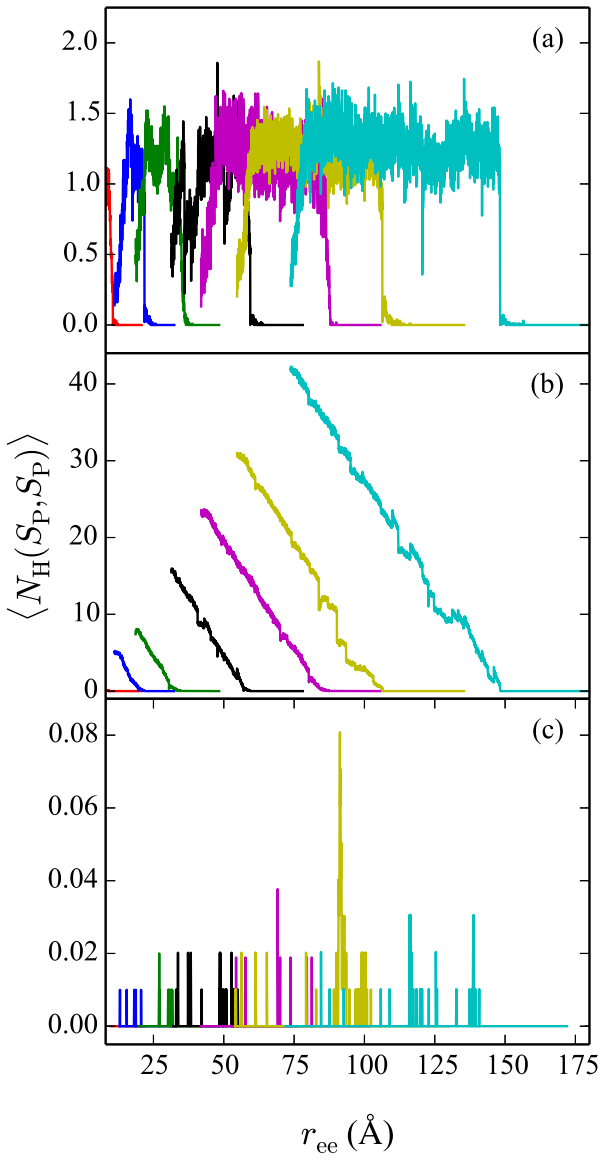

Figure 27: Comparison of the intrapeptide hydrogen bond patterns of seven alanine-rich peptides. Panel (a) shows the $i \to i+3$ bonds, panel (b) shows the $i \to i+4$ bonds, and panel (c) shows the $i \to i+5$ bonds. Each curve is the weighted average obtained using explicit water solvent and 100 tps at 1 Å/ns. The color scheme of the curves is consistent with Fig. 26.

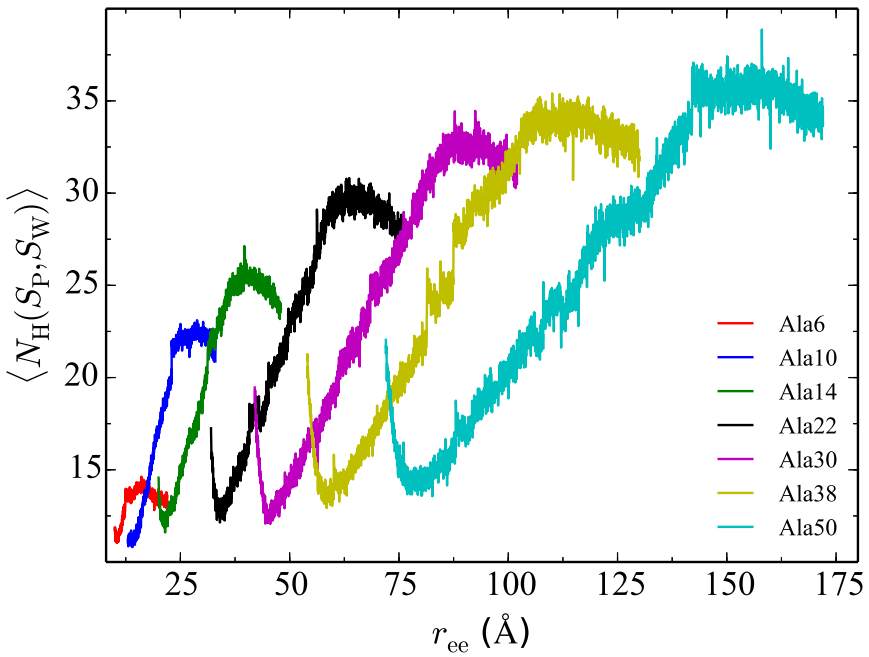

Figure 28: Comparison of the protein-water hydrogen bond patterns of seven alanine-rich peptides. Each curve is the weighted average obtained using explicit water solvent at 100 tps at 1 Å/ns.

The hydrogen bonds formed between the seven alanine-rich peptides and the explicit water solvent are presented in Fig. 28. As in Fig. 27, all of the curves are the weighted averages of 100 trajectories per stage using the averaging scheme described in Chapter 2 and the curves have not been shifted by their initial value of r_{ee} (as the PMFs in Fig. 26) for clarity. Each peptide exhibits a similar trend in the increase of the number of peptide-water bonds formed over the course of the unfolding. Similar to the finding that the number of $i \rightarrow i+4$ hydrogen bonds increases linearly with the length of alanine-rich peptide as shown in Fig. 27b, the number of peptide-water hydrogen bonds increases linearly with the length of the alanine-rich peptides.

3.4 Conclusions

3.4.1 Helical unfolding pathways in implicit solvent

This work provides a comparison of the potentials of mean force and hydrogen-bonding profiles of the forced unfolding of ALA_{10} using primarily two different methods, ASMD and FR-ASMD, at different pulling velocities and in three different environments. Using a variety of velocities and environments, albeit on a simple single-motif peptide, benchmarks for energetics and structure in small molecule pulling simulations have been obtained.

Perhaps surprisingly, there is qualitative agreement between the potentials of mean force and the overall hydrogen-bonding pathway for ALA_{10} stretching in implicit and explicit solvent. However, the equilibrium structure of ALA_{10} in implicit and explicit solvent differ only slightly with the implicit solvent giving rise to a mix of π-helical and α-helical structure while the explicit solvent gives rise to primarily α-helical structure. This suggests that the implicit solvent model slightly over-emphasizes contact between the solvent and peptide. Meanwhile, the fact that the pathways in the presence of an implicit or explicit solvent differs dramatically from that in vacuum illustrates the importance of the use of solvent models (either explicit or implicit) so as to capture the actual dynamics of the stretching. A detailed view of the effective peptide-water hydrogen-bonding (by way of solvating explicit water to the fixed protein structures) in an implicit solvent has also shown strong agreement to the actual peptide-water hydrogen-bonding in an explicit solvent. This observation provides further evidence that hydrogen-bonding is effectively encoded in the implicit model. Thus, the implicit models are capturing the stabilization of the structure with respect to hydrogen bonds in a manner that is analogous to the explicit water. As such, the PMFs obtained from the implicit and explicit water models were qualitatively and quantitatively similar.

Finally, ASMD has been seen to produce a converged PMF of ALA_{10} with fewer trajectories at comparably faster velocities than traditional SMD, and does best in the vacuum case. The use of FR-ASMD, which includes intermediate stages for relaxation without using any type of trajectory selection, does improve the PMF substantially in comparison with naïve ASMD at equivalent pulling velocities in the implicit and explicit solvent cases. This suggests a need for improved trajectory selection schemes in ASMD to address the increased

complexity of the ensemble space in the presence of solvent. Improved contraction schemes that avoid the expense of the propagation of trajectories while not being limited to the limitations of a single representative of the non-equilibrium ensemble are therefore of strong interest [61].

3.4.2 Benchmarking of helical motifs using CHARMM force fields

Molecular simulation promises atomic-detail of key processes in chemistry and biology, but only if the underlying energy potentials and methodology have demonstrated accuracy. Similarly, evaluation of computational molecular mechanisms to explain helical stability of peptides is difficult to postulate due to lack of corresponding experimental data at the atomic level. How helical peptide structure, peptide amphililicity (or lack there of), and specific amino acid sequence influence the formation and stability of secondary structures is still not completely clear. The results presented herein provide an interesting comparison of two very different α-helical peptides, 1PEF and our model system ALA_{10}, across the CHARMM family of molecular dynamic potentials utilizing our ASMD methodology. Our results show the transition between α-helical and random coil states for these linear peptides follow similar pathways of mechanical unfolding across the family of potentials; despite their sequence differences, one is a homopolymer and one is amphiphilic.

For longer peptides, such as 1PEF, the free energy landscape can be very complex with large barriers between different minima. This can be problematic for methodologies that use a single constant temperature molecular dynamics simulation trajectory based on Boltzmann sampling of conformations. These simulations frequently get trapped in one of the local minima and lead to a biased sampling of configurations. Furthermore, the time scale accessible to most computer simulations is often significantly shorter that the biological system's natural time scale. The method of SMD applies an external force to manipulate the peptide in order to probe stability of the peptide along a predefined pulling coordinate, in addition to accelerate the unfolding process that would be too slow to model otherwise and still requires large computational resources to simulate full trajectories for

meaningful statistically analysis. However, performing the mechanical unfolding simulations in an adaptive manner (ASMD) requires fewer non-equilibrium trajectories. As a result, our ASMD non-equilibrium all-atom trajectories shed light on the relative contributions of context dependent helix-stabilizing characteristics along the mechanical unfolding pathway. Our ASMD methodology is robust enough to capture expected behavior across the CHARMM family of FFs; such as the presence of π-helical artifacts in the c22 FF, and the bias of the c27 FF toward α-helical secondary structures. Furthermore, we show that even though 1PEF is an 18-mer, and is comprised of about twice as many amino acids as our model system ALA_{10}, convergence can be reached using 100 tps with a pulling velocity of 1 Å/ns. These thermodynamic results, in combination with the detailed structural and electrostatic analysis as a function of r_{ee}, shows that while 1PEF has a well defined forced pathway of unfolding, it remarkably resembles the unfolding pathway of the homopolymer ALA_{10}. The thermodynamic results of ALA_{10} and 1PEF suggest a concerted unfolding of the hydrogen bond contacts, and transient intermediate structures that have mostly 3_{10}-helical characteristics and hydrogen bonds [153]. This unfolding behavior helps stabilize secondary structures relative to the random coil, and plays a key role in helix stability.

3.4.3 Comparing the performance of CHARMM36 and Amber16 using helical peptides

The PMFs, residue pair interaction energies, and hydrogen bonding patterns of two helical peptides, 1PEF and $gp41_{659-671}$ have been calculated and compared for two different FFs: c36 and Amber16. This works provides a clear and consistent comparison of the performance of two of the most popular potentials used for biomolecular simulations. Our findings suggest that the use of the different potentials does not significantly effect the unfolding energetics and general secondary structure obtained for the peptides. The magnitude and curvature of the PMFs are consistent between the two FFs. 1PEF presents as dominantly an α-helix in both Amber16 and c36 while the structure of $gp41_{659-671}$ is captured as a mix of α- and 3_{10}-helix. However, there are some unexpected and substantial differences in the additional observables that are calculated. The residue pair interaction energies calculated for both 1PEF and $gp41_{659-671}$ show differing treatment of the interactions.

This is surprising given the similarities observed in the PMFs. As such, this may suggest that the are only slightly effecting the overall unfolding pathway of the helical peptide motif. Though it should be noted that this does not discount the utility of salt bridging interactions in any native function or contributing to the solubility of the peptides. This work is further discussed in Chapter 4, using trpzip1 as a benchmark β-hairpin peptide.

3.4.4 Linearity of elongating alanine-rich peptides

In this work, we have elucidated the unfolding pathways of several alanine-rich peptides in explicit water solvent. The PMFs, intrapeptide and peptide-water hydrogen bonding patterns were calculated using ASMD and compared for each peptide. The forced unfolded pathways yielded PMFs consisting of similar curvature for each peptide. The PMFs also show that the addition of a helical turn results in an additive property wherein the PMF is increased. Similarly, the intrapeptide hydrogen bonding patterns showed that each peptide is composed of predominantly α-helical contacts and the peptides lose each type of hydrogen bonding at the same rate. The hydrogen bonding patterns, both intrapeptide and peptide-water, increase linearly with a increase in the number of helical residues. Thus far, these results suggest that the helix-to-coil transition is only dependent on the breaking of the backbone hydrogen bonds and independent of specific side chain interactions and chain length. Overall, this work provides a unique comparison between alanine-rich peptides.

CHAPTER IV

ENERGETICS OF UNFOLDING β-HAIRPIN PEPTIDES

4.1 Introduction

4.1.1 Trpzip1 versus Chignolin

The energetics of unfolding small peptides has become a very useful figure of merit in developing and benchmarking new methods for studying the dynamics of biomolecules [156]. Specifically, the response of peptides or proteins to mechanical folding has been seen to provide insight into their energetics, dynamics, and stability in various complex environments [156]. Small peptides, in particular, are attractive for computational studies because the system sizes are quite manageable on even the most modest of resources while still providing ample insight into the behavior of larger proteins [157]. For example, the helix to random coil transition in polypeptide chains involves a significant change in secondary structure but the simple model introduced by Zimm and Bragg revealed that the $i \rightarrow i+4$ contacts were sufficient to drive the transition as early as 1959 [158]. The determination of the structure and dynamics of more general protein motifs with specificity remains a significant challenge. One such target is the C-terminal fragment of the B1 domain of protein G. After fluorescence spectroscopy experiments by Eaton and coworkers [159] established that it exhibits two-state folding kinetics on the microsecond timescale, the β-hairpin became a model motif that is both experimentally accessible and a good target for computational advances [160].

There are two main theories concerning the driving force behind β-sheet folding: hydrophobic collapse and turn-initiated folding [152, 159, 161, 162, 163, 164]. Dinner et al. [162] investigated the structure and stability of a 16-residue β-hairpin using multicanonical Monte Carlo simulations. They obtained the free energy of folding the peptide along the domain of the number of native peptide bonds. They concluded that the hairpin "zips up" via hydrophobic collapse of tyrosine and phenylalanine residues. However, experimental results

from Eaton and coworkers [159] and computational support from Kolinski, Ilkowski, and Skolnick [161] on the model β-hairpin of the C-terminal fragment of the B1 domain of protein G, indicate a tendency for β-hairpins to fold via turn-initiated folding with only a small influence from hydrophobic collapse. Similarly, Kim and Keiderling [165] used long-time molecular dynamics simulations and principal component analysis to obtain the unfolding pathway of trpzip2 by monitoring the hydrogen bonds over the course of the simulation. They determined that the overall folding mechanism is dominated by the hydrophobic side-chain interactions of the four Trp residues, which support hydrogen bonding contacts. In this work, we further investigate the stability and structure of the β-hairpin motif by obtaining the energetics, hydrogen bond profiles, and specific residue pair interactions of two β-sheet peptides, tryptophan zipper 1 (trpzip1) [166] and chignolin [167], along a pulling pathway using Adaptive Steered Molecular Dynamics (ASMD) [54, 55]. This method is based on the forced unfolding of a peptide through Steered Molecular Dynamics developed by Schulten and coworkers [23] but performed in stages so as to obtain more efficient convergence.

A recurring challenge to the choice of small peptide models (that reduce the complexity of corresponding larger proteins) is the solubility criterion as they are not experimentally accessible if they are not soluble. Fortunately, there have been several experimental designs that have produced soluble and stable monomeric peptides under 20 residues in length [166, 167, 168, 169]. Of all the possible peptide motifs, β-sheets have been seen to be the most appropriate for designing stable peptides that fold independently. There has been much success using tryptophans as stabilizing residues in the design process. Andersen and coworkers designed very stable miniproteins known as "trp cages" [169]. The tryptophan zipper, used in the present work, is another example of a tryptophan-rich miniprotein [166]. Despite being an artificially designed protein, it was recently shown that trpzip3, the 16-residue relative of trpzip1, can inhibit amyloid aggregation of two different amyloid systems: transthyretin and Aβ [170]. The Trp-Trp pairs within trpzip2 and trpzip4 have also been seen to stabilize the structure during mechanical unfolding using detailed single-molecule optical tweezers experiments [171].

Due to the high stability and activation energy of unfolding, it is not practical to use standard MD techniques to simulate the unfolding of most peptides. Advanced numerical techniques such as umbrella sampling [36], replica exchange molecular dynamics (REMD) [80, 172], accelerated MD [173, 174], steered molecular dynamics (SMD) [20], SMD with a hybrid differential relaxation algorithm [175], and diffusion maps [176, 177] have been developed to increase the observation of rare events, such as unfolding. These approaches have also been used to identify the presence of important contacts along the pathway as we do in the current work. Examples of such use include: The free energy profile of the backbone dihedrals in chignolin, alanine dipeptide and a Trp-cage was calculated and compared using several different reweighing methods (exponential average, Maclaurin series, and cumulant expansion) for dihedral-boost and dual-boost accelerated MD simulations [174]. Settani et al. [178] used SMD to determine the effects of ligand binding and to correlate peaks in the force profile to specific residue stabilization. A recent REMD folding study [179] of several β-sheet motifs, including WT and mutants of trpzips 1-3 and B1 domain of protein G, ranked the peptides in terms of stability and suggested that it results from the degree of turn formation. Folding studies of trpzips 1-3 using a combination of aggregated MD simulations and laser T-jump spectroscopy [180] suggest a folding pathway in which the inner Trp pairs, namely Trp4-Trp9, and inner hydrogen bonds form first. Settani and Fersht [181] determined that the rate-limiting step for the unfolding of trpzip1 is the disruption of the hydrophobic core created by the tryptophans using high temperature (373K to 470K) MD simulations of the unfolding of trpzip1 for a total of 6 microseconds. Wales and coworkers [182] compare the rigidity of trpzip1 and chignolin using a local rigid body framework for coarse graining the peptides. In the current work, we aim to obtain the energetics and structural features along the pathway as a protein is stretched apart. The literature precedents cited above suggest that SMD is an appropriate method to use[23, 24, 20, 22] as it offers a direct route for obtaining the energetic and mechanical response of the protein to external driving forces. Specifically, we find that comparison between trpzip1 and chignolin as they are pulled apart using variants of SMD provides a more nuanced perspective on the roles of the individual pair contacts in providing stability along the

pathway, and the differences between the unfolding pathways as a result of an exquisite dependence on the primary structure.

4.1.2 Trpzip1 mutations

The stability of an isolated biomolecule is dictated by numerous factors including amino acid sequence, specific side chain interactions, hydrogen bonding effects, hydrophobic packing, backbone strain, and solvent environment. It is the interplay of these factors which is important; but, it is often difficult to obtain a quantitative understanding of the combination of those effects. Computational studies, particularly molecular dynamics (MD), lend a different perspective to this challenging problem because noncovalent intermolecular interactions can be readily observed and analyzed on the single-molecule scale.

The β-hairpin motif has often been used as a model biomolecular system for both computational and experimental biophysical studies that probe those effects [183]. They are suitable model systems because they are usually small, adopt a specific native fold, and are stable. An isolated peptide motif, such as the β-hairpin, provides a system in which the local environment, such as solvent, can be specifically controlled because the system sizes remain small enough to model computationally but large enough to be of general interest to the community. A family of small, stable β-hairpin peptides, referred to as tryptophan zippers (trpzips) were designed by Starovasnik and coworkers [166]. The smallest of those peptides, trpzip1 (PDB: 1LE0), is the basis for this study. The design rationale behind the trpzips was to design the smallest hairpin structure that was stabilized without the addition of disulfide linkages. There have been several studies involving the various factors controlling the behavior and stability of the trpzips [184]. In addition, trpzip3 has been seen to possess anti-aggregation properties when introduced to two different amyloid-β systems [170]. The first peptide in this family, trpzip1, will be the focus of this work.

Some of the strongest noncovalent intermolecular interactions within peptides are electrostatic in nature. Salt bridges are electrostatic interactions between oppositely charged

residues. In the β-hairpin motif, cross-strand salt bridges have been seen to provide stabilization of the folding pathway and the native structure [185, 186]. However, this stabilization is very sensitive to the system of study and is context and environment dependent. For example, in a study involving several different β-hairpins of various lengths, Chen and coworkers [187] conclude that Glu-Lys pairs are more favorable than Asp-Lys pairs due to side chain length matching. They found that the lengthening by one carbon atom (as in Glu) adds significant electron donating character to carboxylate group by decreasing the withdrawing character of the backbone; creating an enhanced electrostatic interaction. However, a study by Kiehna and Waters [188] determined that the salt bridge contribution was negligible when compared to Phe-Phe interactions. Hydrogen bonds have been previously known as important interactions within peptides [189]. It is well known that hydrogen bonds contribute to the stability of both parallel and anti-parallel β-sheet structures [190]. Hydrogen bonds also contribute to intermolecular interactions between β-sheets such as aggregation behavior and higher-level protein structure formation [190]. Zhao and Wu performed several theoretical studies on the cooperativity of hydrogen bond formation in both helices [191] and β-hairpins [192]. They determined that for β-hairpins, hydrogen bonds do not have cooperative enthalpic contributions. Residues that are generally classified as hydrophobic, such as tryptophan, are known to form stabilized "pockets", which are a driving force behind protein folding. In a mutation study by Sauer and coworkers [193], a buried salt bridge triad within the Arc repressor protein, known to play a critical role in of Arc repressor function, were mutated to hydrophobic residues. Interestingly, the hydrophobic mutant exhibited enhanced stabilization, implying that the the charged residues were not essential for a stable conformation. Maynard, et al. [194] performed thermodynamic analysis by using a combination of NMR and Circular Dichroism methods on the folding of a designed 16-residue analogue of the *met* repressor protein dimer. They concluded that β-hairpin folding is driven by hydrophobic effects, not hydrogen bonding as had been previously suggested. This is in contrast to a trpzip mutation study by Keiderling and coworkers [184], in which they conclude turn stability and hydrophobic packing are the main forces controlling the stability of β-hairpins.

Sequence specific interactions are a significant driving force in peptide stability. Ionizable residues in hydrophobic environments have previously been shown to play a significant role in the dynamics of peptides particularly with respect to the solvation around titratable residue pairs [195]. Previous work [196] on a series of mutants of the B1 fragment of protein G suggest the necessity of "properly placed Coulombic interactions", particularly near the termini. Favorable turn sequences have long been a topic of debate in the β-hairpin community [194]. There is also evidence [197, 198] suggesting that peptide backbone strain plays a role in a number of biological process such as binding, unfolding, and transport.

A molecular dynamics mutation study of the balance between backbone hydrogen bonds, side chain interactions, and stability of the turn region of the 16-residue hairpin fragment of protein G concluded that the side chain interactions are the most sensitive to perturbations in the system, while the hydrophobic core was very stable, and the hydrogen bonds were least stable [199]. On the other hand, it has been shown previously [200] that mutations effect the rate of protein aggregation through a combination of changes to hydrophobicity, charge, and secondary structure propensity. They observed that mutations can either increase or decrease aggregation rates. Unlike previous arguments, the authors suggest that the difference in aggregation rates does not result from specific side chain interactions, but because the mutations effect the overall "physicochemical properties of the system". Thorpe and coworkers [198] used a constraint-based model to investigate the formation of nonnative contacts within the mechanical unfolding of several mutated proteins of varying sizes and secondary structures. However, they were unable to discern whether mutations had a substantial effect on their observables. It is surprising how the predominate intermolecular force varies depending on the system being studied.

In this work, we combine experimental and computational techniques to determine and compare the stability of the trpzip1 and six mutations. The mutations have been designed to investigate the role of specific residues in the stability of the secondary structure and energetics and are illustrated in Fig. 30. We use circular dichroism (CD) spectra and thermal unfolding to experimentally determine the structure and stability of the peptides. Computationally, we determine the single-molecule energetics, hydrogen bonding pattern,

and side-chain interactions for the mechanical unfolding of the trpzip1 and six systematic point mutations.

4.1.3 Comparing the performance of CHARMM36 and Amber16 using the β-hairpin motif

The goal of molecular simulations is to be able to accurately treat the dynamics of any type of system. As such, it is important to benchmark the performance of new potentials and methodologies across a variety of test cases. In this work, the performance of the Amber16 and c36 FFs will be compared using the ASMD methodology with trpzip1. This work is a complementary comparison to the helical motifs presented in Chapter 3. The β-hairpin motif has long been a popular motif with which to compare helical peptides. In a study by Mackerell and coworkers [201] they compared the performance of Amber ff99SB, c27 and c36, for the simulation of unfolded fragments of the NTL9 peptide using residues 122 and 617 using replica exchange MD simulations. Their data suggests that the FFs generally agree in the simulation of the unfolded states and correctly capture the transient formation of β-hairpin structure. However, they also detail how the FFs differ in their treatment of the "native-likeness" of the unfolded/semi-folded structures. Another study by Karttunen and coworkers [109] used a 16-residue fragment of the Nrf2 peptide known to form β-hairpins using 10 different MD potentials: Amber ff99SB-ILDN, Amber ff99SB*-ILDN, Amber ff99SB, Amber ff99SB*, Amber ff03, Amber ff03*, GROMOS96 43a1p, GROMOS96 53a6, c27, and OPLS-AA FFs.

4.2 β-hairpin model details

The β-hairpin peptides investigated in this Chapter are trpzip1, chignolin, and six mutations of trpzip1. In the following sections, a detailed description of the design of the systems and simulation parameters used will be presented.

4.2.1 Trpzip1 and chignolin

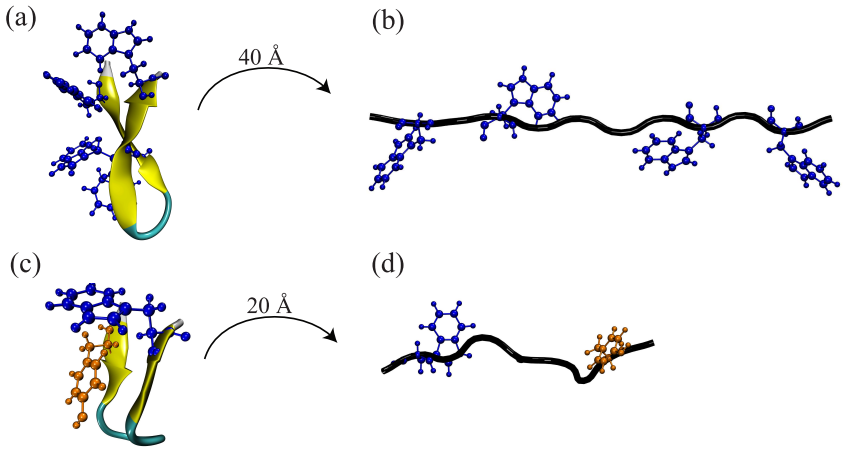

Figure 29: Structures of trpzip1 (top) and chignolin (bottom) are shown in the ribbon representation. The peptides are decorated with atoms from selected residues: The four tryptophans of trpzip1 are highlighted in blue. The single tryptophan of chignolin is highlighted in blue and the tyrosine is highlighted in orange. The initial folded structures of trpzip1 and chignolin are shown at left (a and c, respectively), while the unweighted averaged structures at the end of the ASMD stretches are shown on the right (b and d, respectively). The stretching distances of the peptides are noted above the curved arrows.

The trpzip1 peptide (PDB: 1LE0) [166] belongs to a family of designed peptides known as tryptophan zippers. By construction, they contain multiple tryptophan residues that have favorable aromatic stacking interactions between the strands leading to β-sheet structure of 12 residues with the sequence SWTWEGNKWTWK. They are also monomeric and highly soluble in water. They form extremely-stable well-defined tertiary structures despite not having disulfide bridges, metals or non-standard amino acids.

The chignolin peptide (PDB: 1UAO) [167] consists of 10 residues with the sequence GYDPETGTWG. Like trpzip1, chignolin is a small, stable β-hairpin. Chignolin was designed to have a secondary structure similar to residues 42-52 of the B1 fragment of protein G, which has been a target of several investigations [152, 159, 161, 202, 203].

Although chignolin and trpzip1 share similar secondary structures, they have very different known melting temperatures, 39 °C and 51 °C, respectively. The melting temperature difference is indicative of significant differences in stability leading to detectable differences in their response to mechanical unfolding. which we find here to also be the case along the mechanical unfolding pathway.

All MD simulations performed here are propagated using NAMD [139] with the c27 FF [81, 82, 95]. It should be noted that c27 has been shown to over stabilize helical secondary structure. In a study on a three-stranded β-sheet, the WW domain [204], MD simulations using c27 were unable to obtain the correct native folded structure. Although c36 [113, 205] is now available, it was not employed in order to maintain consistency in the comparisons to prior work by other groups on the chosen proteins and peptides. However, as a test, we performed an ASMD simulation of trpzip1 in explicit solvent using c36 with 100 trajectories per stage at a pulling velocity of 1 Å/ns. This PMF was then compared to the PMF obtained using c27. There was no appreciable difference in the energetics obtained between c27 and c36 (results shown in Supporting Information).

For explicit solvent simulations, N_{W} TIP3P water molecules are employed in a square cuboid with two equal sides of length L_{xy}, and a longer side along the z-axis of length L_z. The PDB file of trpzip1 [166] contains an ensemble of twenty NMR structures. These twenty NMR structures were individually equilibrated in vacuum for 1 ns using NVT conditions and fixed end-to-end length. The latter constraint, both in vacuum and explicit conditions, is imposed by constraining the relative positions of the Cα atoms of the first and twelfth residues. The tenth structure of the trpzip1 PDB file was chosen for this work as it was the best at maintaining the β-sheet motif during equilibration with the chosen force field. This result was verified using the NAMD plug-in Timeline for analyzing secondary structure and determining root mean squared deviations (RMSDs). For the explicit solvent simulations, the initial solvated peptide configuration is obtained by insertion of the vacuum structure in an equilibrated water box with the water molecules overlapping with the peptide removed, and subsequent equilibration for 1 ns with fixed end-to-end peptide length but all other motions unconstrained. For trpzip1, N_{W}=1421, L_{xy}=40 Å, and L_z=60 Å. The stability of

the β-sheet motif in the solvated structures was confirmed by secondary structure analysis after equilibration. The peptide is subsequently stretched along the long (z) axis during the simulation. All simulations are carried out at a temperature of 300 K. The stretching coordinate is defined as the distance between the Cα of the first and twelfth residues. At the start of each simulation, the distance between the stationary and pulled atom is constrained to 4 Å. This constraint ensures that the peptide accesses the minimum of the PMF (which lies at approximately 4.7 Å) during the stretching simulation. Chignolin is prepared under similar conditions to trpzip1. The explicit solvent parameters are N_{W}=1253, L_{xy}=30 Å, and L_z=40 Å. The stretching coordinate is defined as the distance between the Cα of the first and tenth residues. The initial distance constraint between the stationary and pulled atom is fixed to 4 Å. This construction ensures that the minimum of the PMF curve will lie to the right of the initial point as displayed in Fig. 1. Given this point, the PMFs can then be shifted to reset the zero of energy to the minimum. The remaining figures —Figs. 31-34— are shifted accordingly.

The PMF of trpzip1 is evaluated at three different stretching velocities, v_{s}, in vacuum (100, 10, and 1 Å/ns) and four different v_{s} in explicit solvent (100, 10, 1, and 0.1 Å/ns). The trpzip1 peptide is stretched for 40 Å, resulting in a domain of the end-to-end distance from 4 Å to 44 Å. There were several sets of varying sampling sizes, from 10 to 800 trajectories per stage, evaluated for the vacuum and explicit cases. The PMF of chignolin in explicit solvent is obtained at three different v_{s} (10, 1 and 0.1 Å/ns) using the same range of sampling sizes. Chignolin is stretched 20 Å, resulting in a domain of the r_{ee} from 4 Å to 24 Å. The ASMD simulations use twenty and ten evenly partitioned stages along the stretching coordinate— viz. the end-to-end distance, r_{ee}, —in the simulations for trpzip1 and chignolin, respectively. In both cases, the peptides are pulled for 2 Å within each stage.

4.2.2 Trpzip1 mutations

4.2.2.1 Computational methods

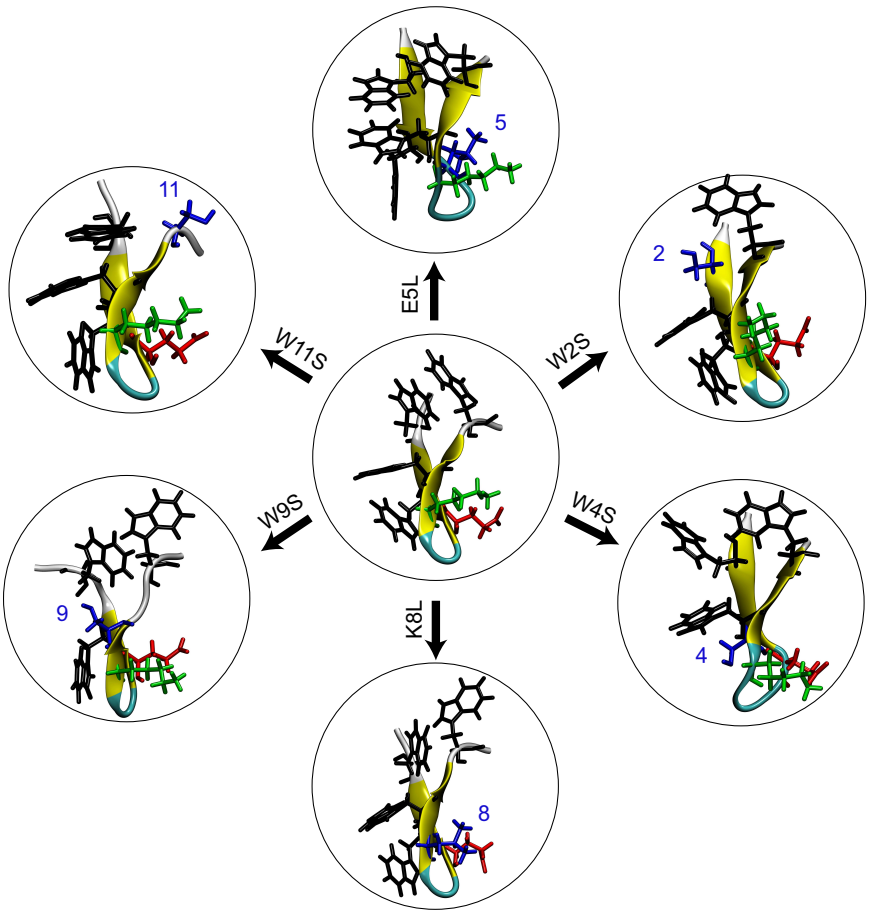

Figure 30: An illustration of the seven peptides studied in this work. At the center is the WT trpzip1. The six peptides around the WT are the mutations and are labeled over the arrow. On the WT peptide, the residue side chains in black, green, and red are the tryptophans (at positions 2, 4, 9, and 11), lysine-8, and glutamic acid-5, respectively. In the mutant peptides surrounding the WT peptide, the mutated residue side chains are shown in blue. All structures shown are the equilibrated structures.

The initial primary sequence of the wild type (WT) trpzip1 is: SWTWEGNKWTWK. For the simulations, six point mutants were built using the VMD plugin Mutator [63]: E5L, K8L, W2S, W4S, W9S, and W11S. The mutants are denoted using the original residue, residue

number, and new residue. The equilibrated WT trpzip1 structure was used as the base for the mutations. The specific mutations were chosen in order to probe the hydrophobic core and the Glu-Lys salt bridge. Leucine was chosen as the replacement residue for the salt bridge residues because of its propensity as a β-sheet forming amino acid [206, 207, 208]. The mutations did not cause any significant structural changes in the backbone of the secondary structure, such that each mutant forms and maintains a stable β-hairpin. This was observed through use of the VMD plugin Timeline for secondary structure analysis of the equilibration of each peptide.

All molecular dynamics simulations of the peptides discussed in Sec. 4.3.2.1 were performed using NAMD [139] with the c27 FF [81, 82, 95]. Several point mutations of the WT trpzip1 were performed. After each mutation, the peptides were equilibrated in vacuum for one nanosecond. Each peptide was solvated using approximately $N_{\mathrm{W}} \approx$5700 TIP3P water molecules in a square cuboid solvent box with two equal sides of length $L_{xy} \approx$47 Å, and a longer side along the z-axis of length $L_z \approx$82 Å. The systems were then ionized for neutrality using Na or Cl ions as implemented in the NAMD protocol for ionization. In E5L, three Cl ions are added, in K8L there is one Cl ion, and in the Trp to Ser mutants there are 2 Cl ions.

After ensuring neutrality, the systems are equilibrated for 1 nanosecond in explicit solvent using NVT conditions. In each equilibration procedure, the Cα of the first and twelfth residues are held fixed on the z-axis. After equilibration, the peptides were analyzed using the NAMD plug-in Timeline for secondary structure stability and calculation of the root mean squared deviations . Secondary structure analysis of the solvated structure was carried out to ensure the stability of the β-sheet motif after equilibration.

During the ASMD simulations, the peptides are stretched along the z-axis. The reaction coordinate is defined as the distance between the Cα of the first and twelfth residues. The Cα of the first residue is held fixed, while the harmonic steering potential is applied to the twelfth residue. At the start of the first stage of each simulation, the distance between the stationary and pulled atom is constrained to 4 Å. This constraint ensures that the peptide does reach a minimum at $\approx$ 4.7 Å during the pulling simulation. All equilibrations and

simulations are carried out at a temperature of 300 K.

For each peptide, the PMF is obtained using a pulling velocity of 1 Å /ns and a sampling size of 100 trajectories per stage, over the entire ASMD simulation of 20 evenly partitioned stages. The peptides are stretched for 40 Å for a total reaction coordinate, r_{ee}, of 4 to 44 Å.

The Root Mean Square Deviations (RMSDs) are compared for each mutation at four different values of r_{ee}, 12, 16, 24, and 28 Å. These distances were selected based on the PMF curves obtained in Fig. 40. Snapshots of the structures of each mutant at those distances were obtained from an average structure from the individual trajectories. Next, the structures were aligned using the RMSD Calculator Plugin available in the VMD package [63]. The structures were aligned using only the backbone of all of the residues (i.e. 1 through 12). The RMSDs were calculated one at a time using the WT structure and one mutant structure. The RMSD values are reported underneath the corresponding structure illustrated in Fig. 41.

4.2.2.2 Experimental methods

All experiments were carried out by Dr. Stephen Quirk.

Peptides were synthesized using standard solid phase synthetic techniques by New England Peptide, Inc. Each peptide contained a standard NH_2 N-terminal structure and an amidated C-terminal end. Molecular weight and purity were confirmed by mass spectroscopy and reversed phase HPLC. Typical peptide purity was >95%.

Peptides were dissolved to a final concentration of 20 μM in 15 mM $NaPO_4$ (pH 7.0) for far UV (190 to 250 nm) measurements and at 200 μM in the same buffer for near UV (250 to 320 nm) measurements. Spectra were obtained as a function of temperature on an Applied Photophysics Chirascan spectrophotometer utilizing a 1 mm pathlength quartz cuvette. Melting curves were obtained in both UV regions at 1°C intervals after a 5 minute incubation at the new temperature with an averaging time of 5 seconds. Thermal denaturation was fully reversible as evidenced by recovering $\approx$ 98% of the CD signal upon cooling and by the observation that reverse and forward melting curves were superimposable.

Raw CD data was converted to mean molar ellipticity according to:

$$[\theta]_m = \frac{m}{lnC} \tag{9}$$

where m is the raw CD reading in millidegrees, l is the cell path length in millimeters, n is the number of amino acids in the peptide, and C is the micromolar concentration.

CD signal at 228 nm (far UV) or 295 nm (near UV) were fit to a 2-state equilibrium unfolding model after a linear fit of the folded and unfolded baselines according to:

$$F_u = \frac{\theta_f - \theta_i}{\theta_f - \theta_u} \tag{10}$$

where F_u is the fraction folded, θ_f, θ_u, and θ_i are the CD signal of the native, unfolded, and i_{th} temperature state respectively. Finally, the data presented represent the average of three independent experiments, all of which were superimposable to within 0.25°C of T_m. Thermodynamic parameters were extracted from the F_u versus T data by first calculating the equilibrium constant in the form of:

$$K_{eq} = \frac{F_u}{1 - F_u} = \frac{\theta_f - \theta_i}{\theta_i - \theta_u} \tag{11}$$

The Gibbs free energy values were calculated as follows:

$$\Delta G = -RT \ln \left[\frac{\theta_f - \theta_i}{\theta_i - \theta_u} \right] \tag{12}$$

and ΔS_m was determined as the slope of the ΔG versus T plot and is the entropy at the transition temperature (T_m). At T_m, $\Delta G = 0$ therefore $\Delta H_m = T_m \Delta S_m$. All ΔH_m and ΔS_m values were verified using the van't Hoff equation:

$$\frac{d \ln K}{d(\frac{1}{T})} = -\frac{\Delta H}{R} \tag{13}$$

Finally, $\Delta(\Delta G)$ values were calculated thermodynamically for the mutants via the relation of Becktel and Schellman:[209]

$$\Delta(\Delta G) = \Delta T_m \Delta S_m = \Delta T_m \left(\frac{\Delta H_m}{T_m} \right) \tag{14}$$

where ΔS_m and ΔH_m values are for the wild-type peptide. Finally all thermodynamic values were compared to a global fit of all CD data between 200 and 250 nm (Far UV) or between 260 and 315 nm (Near UV) using software provided by Applied Photophysics with the Chirascan CD unit. One peptide in the study, E5L, does not follow a two state unfolding mechanism. For that peptide, the data are fit to an unfolding model with a single intermediate: N $\leftrightarrow$ I $\leftrightarrow$ D.

4.3 Results and Discussion

4.3.1 Trpzip versus chignolin

4.3.1.1 Energetics

As shown by Bryant, Pande, and Rokhsar [152], the choice of pulling velocity affects the overall energetics and pathway that can be obtained from mechanical unfolding simulations. Therefore in this work, for each peptide, convergence is shown for a range of pulling velocities and sampling sizes. The convergence of the PMF for trpzip1 in vacuum with respect to v_s and number of trajectories is shown in Fig. 31. There is a drop in the magnitude of the PMF from 100 to 10 Å/ns, but the values change little with a subsequent drop to 1 Å/ns, suggesting convergence at circa 10 Å/ns. The PMF at 10 Å/ns converges to that as seen in the bottom panel of Fig. 31b. The increases in sampling size for the 100 and 10 Å/ns stretching velocities leads to little or no change in the PMFs, hence demonstrating convergence. The convergence behavior seen here is similar to that for our earlier simulations on ALA_{10} in vacuum [56].

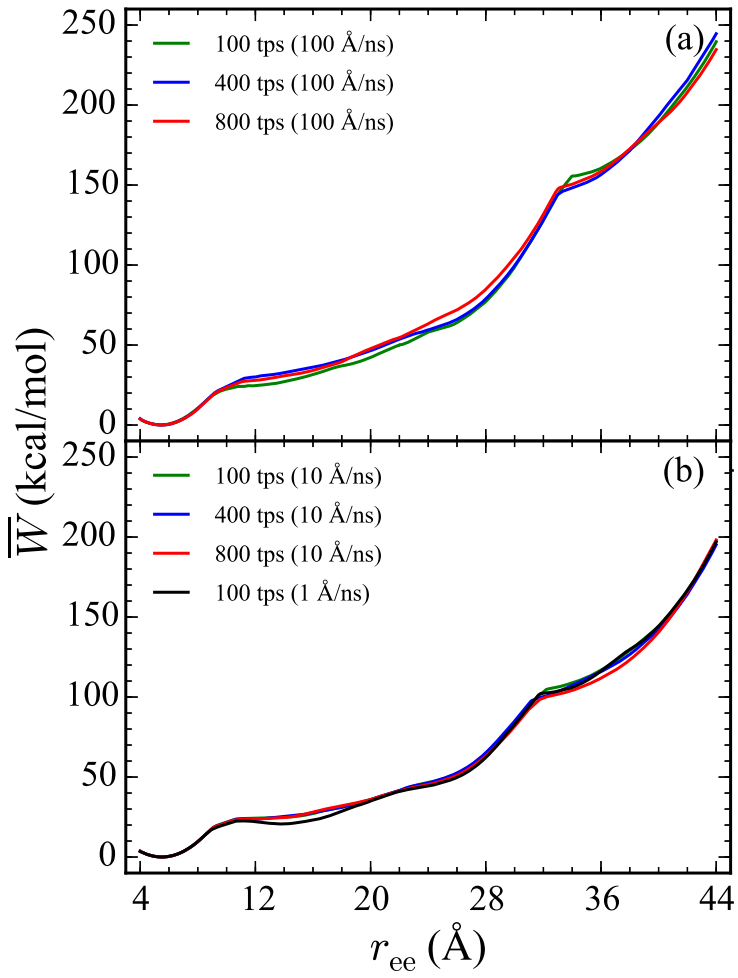

Figure 31: PMFs of trpzip1 calculated in vacuum using three different stretching velocities v_s. Panel (a) and (b) correspond to 100 Å/ns and 10 Å/ns, respectively, except for the additional PMF in panel (b) shown in black corresponding to 1 Å/ns. In both panels, red, blue, and green curves correspond to 800, 400, and 100 trajectories per stage (tps).

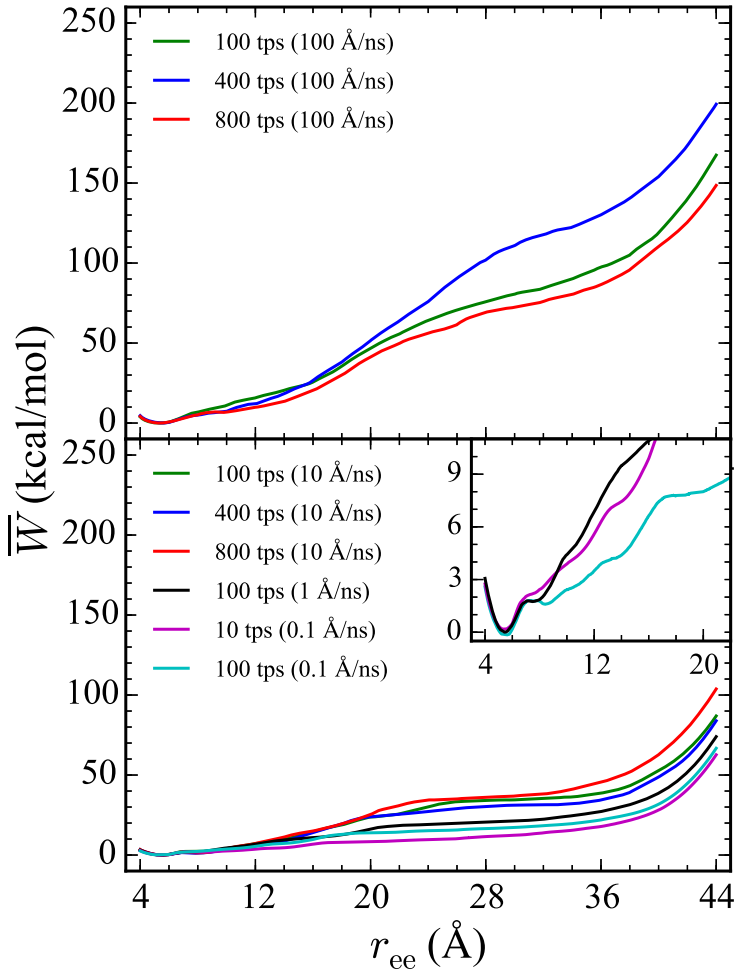

Figure 32: PMFs of trpzip1 calculated in explicit solvent at four different values of v_s. In both the top and bottom panels, the results are represented using the same color scheme and velocities as shown in Fig. 31, with the new additions in panel (b) for 10 tps at 1 Å/ns (purple) and 100 tps at .1 Å/ns (cyan). The inset in panel (b) reveals the PMF between 4 and 20 Å.

In explicit solvent, the trpzip1 PMF was calculated for v_s of 100, 10, 1, and 0.1 Å/ns at several sampling sizes as shown in Fig. 32. Although the most converged data is obtained at 0.1 Å/ns with a sampling size of 100 trajectories per stage, as shown in the cyan curve, we are able to obtain convergence of the PMF at a v_s of circa 1 Å/ns as also illustrated in

the bottom panel of Fig. 32. The PMF for the trpzip1 peptide in explicit solvent compared to that for vacuum has decreased by approximately 100 kcal/mol by the end of the stretch. At a v_s of 0.1 Å/ns, the increase in sampling from 10 to 100 trajectories per stage slightly improves convergence of the PMF as shown in the inset of Fig. 32.

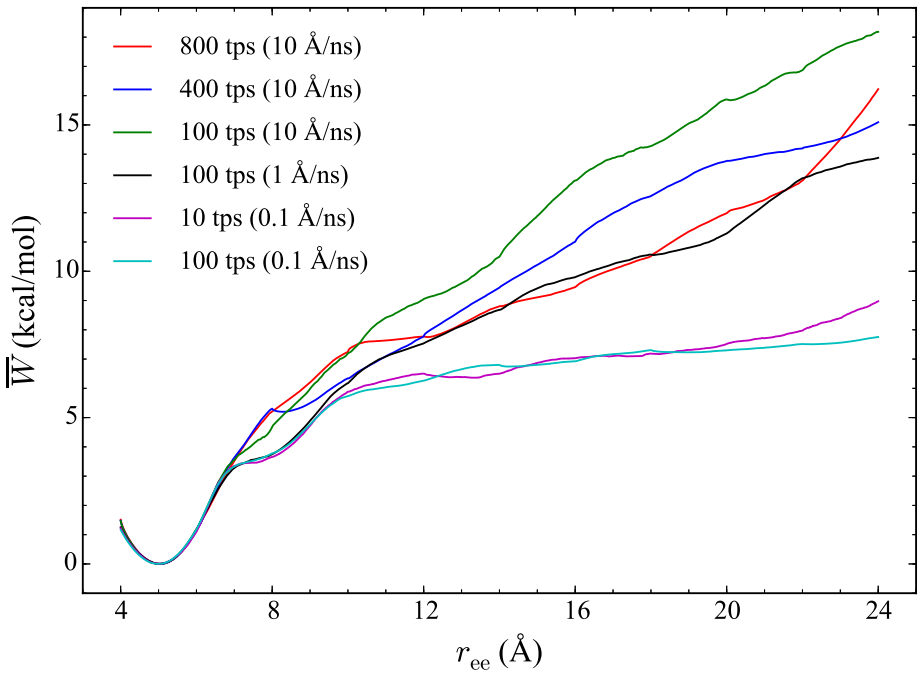

Figure 33: PMFs of chignolin calculated using ASMD in explicit solvent at three different values of v_s. The PMFs in red, blue, and green correspond to 800, 400, and 100 trajectories per stage (tps) obtained at v_s =10 Å/ns. The black curve shows the PMF of 100 tps at 1 Å/ns. The magenta and cyan curves show two separate runs of 10 tps and 100 tps at 0.1 Å/ns.

The PMF of chignolin was calculated in explicit solvent at 3 different v_s: 10, 1, and 0.1 Å/ns as shown in Fig. 33. Although trpzip1 is only longer than chignolin by two residues, the full stretch of trpzip1 requires 40 Å, while chignolin requires 20 Å in order to obtain coils without significant secondary contacts. This is also reflected in the energetics in so far

as the bulk of the work required to stretch trpzip1 takes place in the first 10 Å, whereas chignolin needed only 5 Å. Thus the energetic behavior suggests that chignolin is a less stable peptide than trpzip1 despite having similar structure and characteristics.

As shown in Fig. 31, the PMF of the forced unfolding of trpzip1 converges very rapidly in vacuum. As the velocity decreases from 100 to 10 Å/ns, the final value of the PMF decreases by roughly 60 kcal/mol. When the velocity in decreased by another factor of 10 to 1 Å/ns, there is no decrease in the PMF, as shown in the black curve in Fig. 31b. The PMF is approximately 100 kcal lower in explicit solvent and the PMF becomes smoother due to interaction with the solvent. In Fig. 34, trpzip1 at 0.1 Å/ns with 100 trajectories per stage (the green curve) exhibits a slight minimum at approximately 8.5 Å. This is also confirmed by a short plateau of the hydrogen bonding at 3 hydrogen bonds as shown in Fig. 35b in the 8 to 12 Å range. This may suggest the presence of a stable intermediate, though one was not directly observed. Previous work uncovering the unfolding/folding pathways in similar peptides such as the 16-residue C-terminal fragment (41-56) of protein G-B1 [210] and trpzip2 [211] suggest the existence of intermediates stabilized by hydrogen bonds and the hydrophobic core. This is also in agreement with findings by Wang and coworkers [212], in which trpzip2 was found to fold (using REMD) by a mechanism of hydrogen bonding and hydrophobic core collapse.

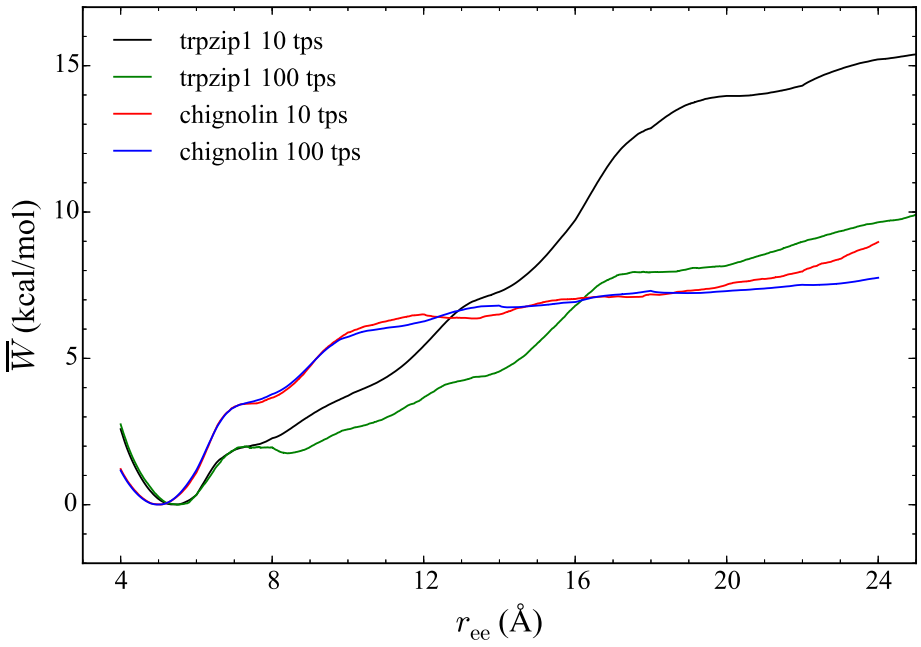

Figure 34: PMFs of chignolin (red and blue) and trpzip1 (black and green) calculated in explicit solvent at 0.1 Å/ns. Curves in black and red correspond to 10 tps, and curves in green and blue correspond to 100 tps.

Also shown in Fig. 34, trpzip1 has a shifted minimum at the beginning of the simulations compared to chignolin. The trpzip1 PMF minimum occurs at a r_{ee} of 5.45 Å , while the minimum PMF value obtained for chignolin occurs at a r_{ee} of 4.93 Å. The PMF of trpzip1 remains lower in energy than chignolin until it reaches a stretching distance r_{ee} of approximately 12-16 Å. After that distance, the chignolin PMF becomes the lower of the two PMFs. The primary significance of these features, as noted above, are that the most of the work required to stretch chignolin occurs over a shorter pull and that more work is required to fully open trpzip1.

The PMFs obtained by ASMD for both β-sheet peptides, trpzip1 and chignolin, in comparison with the earlier stretching energetics [55, 56] of the α-helical ALA_{10}, suggest

that the β-sheet is more stable in vacuum but not necessarily in explicit solvent. Specifically, ALA_{10} in vacuum [55] requires approximately 20 kcal/mol to remove the majority of its structure across 20 Å whereas trpzip1 requires approximately 200 kcal/mol. Trpzip1 is therefore approximately 10 times more stable than ALA_{10} in vacuum. However, ALA_{10} in solvent [56] requires approximately 10 kcal/mol to remove the majority of its structure across 20 Å which is very similar to the PMF obtained for chignolin in explicit solvent. Trpzip1 does display more stabilization than either chignolin or ALA_{10} in explicit solvent. Trpzip1 requires approximately 50 kcal/mol to completely unstructure the peptide. The stability of the β-hairpin thus appears to be very sensitive to differences in sequence and specific residue interactions within the peptide. This reinforces previous findings [159] that β-sheets can display increased stabilization over α-helices. While the two β-hairpins investigated here are of similar length, they have dramatically different energetics.

4.3.1.2 Hydrogen bonds

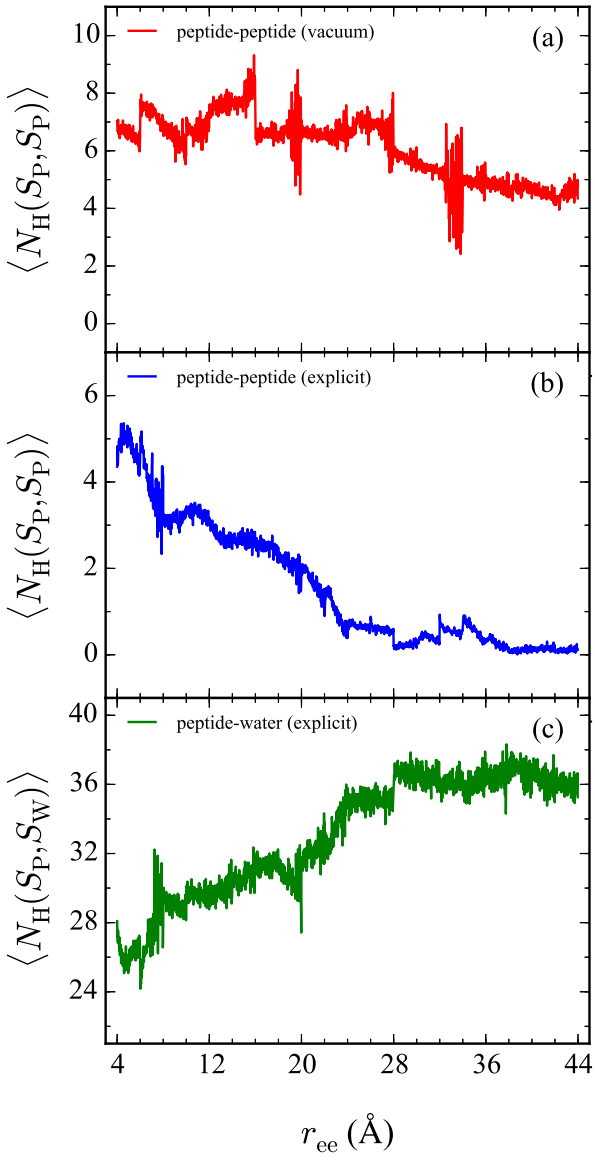

Figure 35: The weighted average number of intrapeptide hydrogen bonds for trpzip1 in two different environments, vacuum (a) and explicit solvent(b) are shown. The weighted average number of hydrogen bonds formed between trpzip1 and the explicit solvent is shown in panel (c). For each panel, the average is performed over 100 tps at $v_s = 1$ Å/ns.

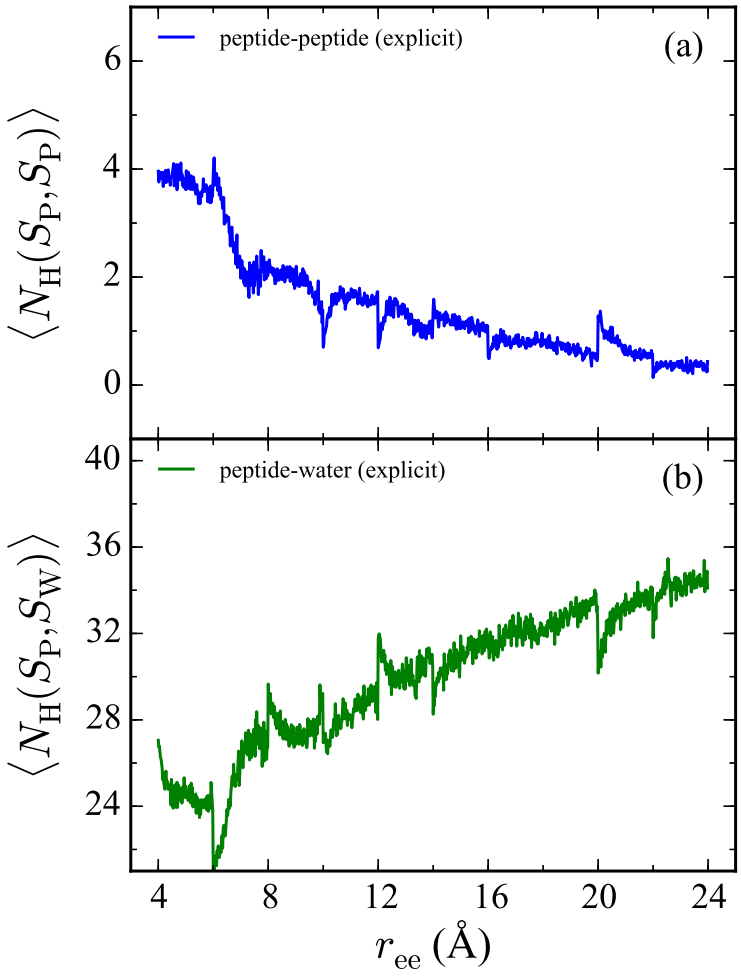

Figure 36: The weighted average number of intrapeptide hydrogen bonds formed in chignolin in explicit solvent in shown in panel (a). The weighted average number of hydrogen bonds formed between chignolin and the explicit solvent in shown in panel (b). For each panel, the average is performed over 100 tps at $v_s = 1$ Å/ns.

The hydrogen bonds maintained during the stretching of trpzip1 in vacuum is one possible reason why the PMF is considerably larger than in explicit solvent. By the end of the unfolding, the peptide maintains approximately four hydrogen bonds as shown in Fig. 35a. In explicit solvent, the number of intrapeptide hydrogen bonds decreases steadily to zero by the end of the stretch as can be seen in Fig. 35b. The number of hydrogen bond contacts

the peptide forms with solvent can also be readily seen over the course of the simulation. As shown in Fig. 35c, the peptide increases the number of bonds to surrounding water molecules to 36. The bonds formed between the peptide and the water over the course of pulling facilitate the unfolding of trpzip1 in solvent. The persistence of the peptide-peptide hydrogen bonds maintaining secondary structure of the β-hairpin that was observed in vacuum is not observed here. Instead, in explicit solvent, the loss of structure in response to forced unfolding is facilitated by the hydrogen bonds formed between the peptide and water. This stabilization of the peptide by the solvent during the stretch creates a possible unfolding pathway accessible to trpzip1 that is much lower in energy than what is available in vacuum conditions.

A similar hydrogen bonding trend in explicit solvent can be observed for chignolin. The number of intrapeptide hydrogen bonds decreases from an average of 4 to 0 in explicit solvent, as shown in Fig. 36a, while the number of hydrogen bonds formed between the peptide and surrounding water molecules increases from an average of 27 to 35, as can be seen in Fig. 36b. Zhao and Cheng [213] used SMD to study the unfolding of the 10-residue chignolin mutant CLN025 [168] (sequence: YYDPETGTWY). This protein has two sets of stacked aromatic residues: Tyr1-Tyr10 and Tyr2-Trp9. Although, unable to obtain a potential of mean force for the unfolding of CLN025, they were able to analyze the hydrogen bonding and hydrophobic interaction between the aromatic pairs. In similar agreement with the present results, they found that CLN025 has approximately 6 intrapeptide hydrogen bonds in the folded structure that decrease to 0 bonds by the end of the pulling.

The hydrogen bonding profiles, although weighted in accordance with the work values, still have significant fluctuations. For the trpzip1 vacuum simulations, there are three specific regions of the hydrogen bond profile that are of interest at end-to-end distances of 20, 28, and 33 Å. In these regions, the fluctuations of the bonds are significant. Those regions also correspond to specific interactions seen in the residue pair interaction energies in Sec. 4.3.1.3 at those particular distances. In comparison with the hydrogen bonding profiles of the small helical peptides obtained previously [55, 56], the β-sheet motif is considerably more stable, particularly in vacuum conditions. In our earlier helical peptide studies, there

are different types of hydrogen bonds (α-, 3_{10}- and π-helix) that are capable of forming along the unfolding pathway that add stabilization during the unfolding. In the case of β-sheet peptides, no such bonds form. However, hydrogen bonding networks, both intrapeptide and solvent-peptide, have been shown [152, 202, 203] to provide significant stabilization of a similar hairpin motif. Notably the hydrogen bonding contacts in the trpzip2 peptide —similar in structure and stability to trpzip1— were found [214] to play a substantial role in supporting and facilitating the hydrophobic core collapse during folding. The present results exhibit stabilization during the analogous mechanical unfolding of both trpzip1 and chignolin and are therefore in agreement with previous findings for hairpins of similar size and structure.

4.3.1.3 Residue pair interactions

As was previously shown [187, 196], the interactions of amino acid side chains play a significant role in stabilizing β-hairpins. In order to further probe the stability of the two peptides in this study, we now evaluate the effective interaction energy between different of pairs of residues along r_{ee}. The interaction energies are obtained using weighted averages according to Eq. 7.

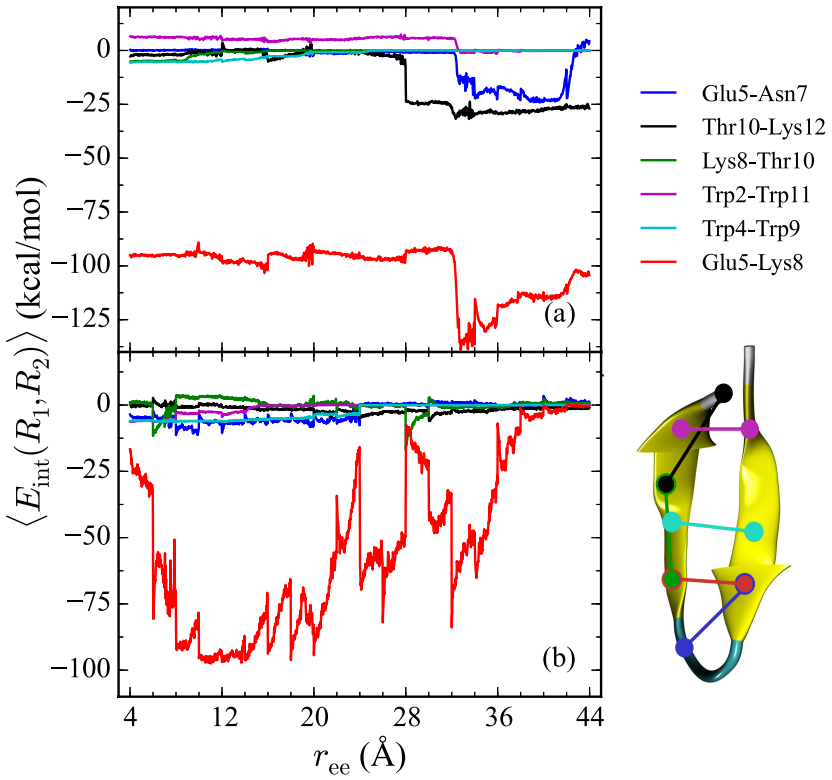

Figure 37: Comparison of several residue pair interaction energies within the trpzip1 in vacuum (a) and in explicit solvent(b). The energies for each curve are obtained by a weighted average of 100 tps at $v_s = 1$ Å/ns. The cartoon shows the approximate locations of the selected residue pairs.

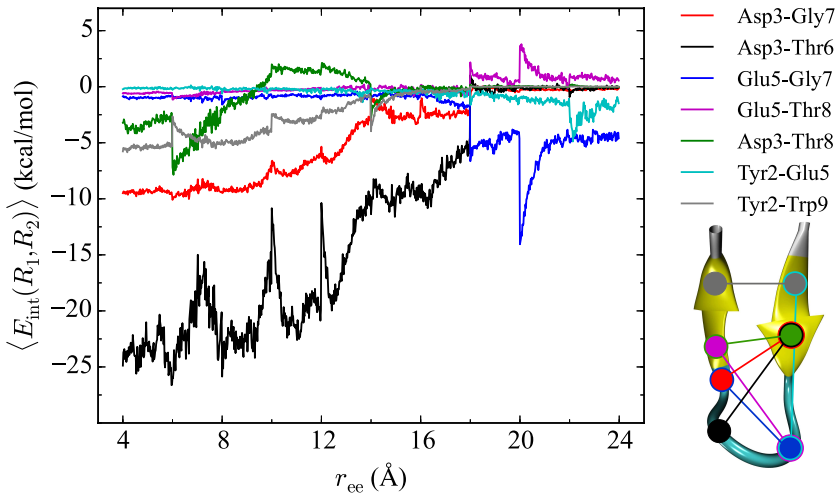

Figure 38: Comparison of several residue pair interaction energies within chignolin in explicit solvent. The energies for each curve are obtained by a weighted average of 100 tps at $v_{\rm s} = 0.1$ Å/ns. The cartoon shows the approximate locations of the selected residue pairs.

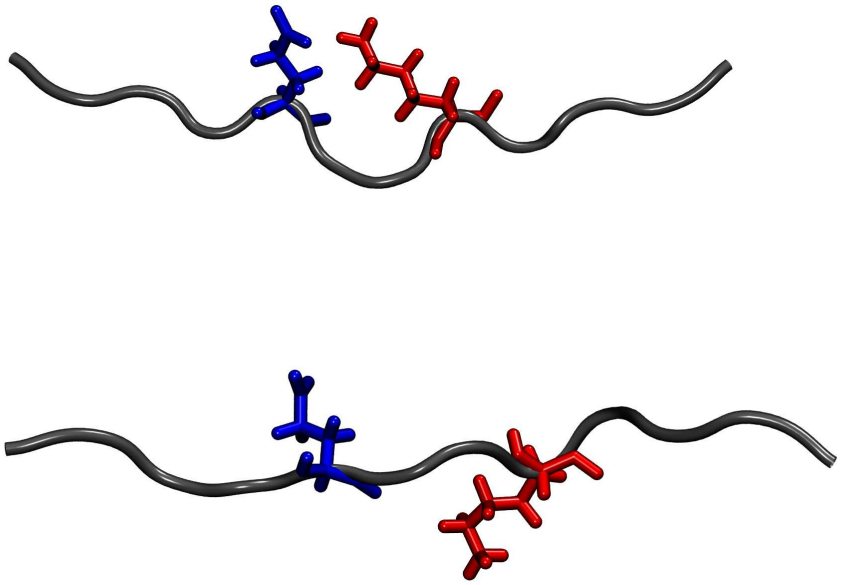

Figure 39: Unweighted average structures of trpzip1 in explicit water are shown to illustrate the loss of the stabilizing interaction between Glu5 (blue) and Lys8 (red). In the top and bottom structures, the Glu5-Lys8 is interacting at an end-to-end distance of $r_{ee} \approx 34.7$ Åand 36.2 Å, respectively.

For trpzip1, several residue pairs were identified and monitored over the course of the unfolding: Glu5-Lys8, Trp2-Trp11, Trp4-Trp9, Lys8-Thr10, Thr10-Lys12, and Glu5-Asn7. These were chosen either because they are involved in a contact in the folded structure, are involved in a salt bridge or aromatic stacking, or are potentially affecting a known contact. There are three residue pairs which exhibit very pronounced stabilization: Glu5-Asn7, Thr10-Lys12, and Glu5-Lys8. Among these pairs, the most stabilizing interaction is the Glu5-Lys8. It forms a salt bridge near the turn region of the peptide as shown in red on the right of Fig. 37. The stabilization of the Glu5-Lys8 residue pair in vacuum, shown in Fig. 37a, is very pronounced and maintained for the entire stretch of the peptide. In explicit solvent, this stabilization is observed for a significant portion of the stretch, but tapers off towards the end of the stretch at $r_{ee} \approx 36$ Å as seen in Fig. 37b. When the salt bridge breaks, the resulting residues are stabilized through contacts with water that are

unavailable to the peptide in vacuum. Example structures before and after this transition are shown in Fig. 39. In the top structure with an end-to-end distance of $r_{ee} \approx 34.7$ Å, Glu5 (blue) and Lys8 (red) are clearly oriented toward each other because of a favorable interaction. In the bottom structure, with an end-to-end distance ($r_{ee} \approx 36.2$ Å) extended nearly 2.0 Å further, the potential salt bridge between those residues is no longer formed. As such, it should not be surprising that the Glu5-Lys8 contact is comparatively more stable in vacuum than in solvent across the entire pull. This is in agreement with observations by Searle and coworkers [185] who induced two Lys-Glu interactions through mutations of a model 16-residue β-hairpin. They concluded that the double mutants were much more stable than the single mutants and that the salt bridge interactions lingered for up to 75% of the simulations. This is also in agreement with previous folding simulations [215] carried out using Markov state models and kinetic networks in which the importance of the Glu5-Lys8 residue pair was noted as a important stabilizing contact.

Similarly, several residue pairs were monitored within chignolin: Asp3-Gly7, Asp3-Thr6, Glu5-Gly7, Glu5-Thr8, Gly1-Asp3, Pro4-Thr6, Thr8-Gly10, Gly7-Trp9, Tyr2-Glu5, and Tyr2-Trp9. Although aromatic stacking plays a role in several of the contacts in both trpzip1 and chignolin —such as the Trp2-Trp11 and Trp4-Trp9 pairs in trpzip1, and Tyr2-Trp9 pair in chignolin— the overall energetics of the residue pairs in chignolin are significantly smaller than those obtained for trpzip1. Asp3, Pro4, Gly7 were chosen for this study because they had been previously been implicated in turn initiation and stabilization. Tyr2 and Trp9 were chosen because they are involved in aromatic stacking interactions [216]. Lovas and coworkers performed MD simulations and used quantum chemical calculations in a variety of solvents and temperatures to determine the interaction energies of several residue pair interactions of CLN025 [217]. As noted previously, in Sec. 4.3.1.2, CLN025 is a mutant of chignolin. In MD simulations, Lovas and coworkers [217] observed that CLN025 maintains a stable β-hairpin conformation over a wide range of conditions: in water solvent at 278, 300, 333, and 363 K and in TFE, MeOH, and DMSO at 300 K, From their quantum mechanical calculations in water at 300 K, they obtained -4.66 kcal/mol in stabilization for the aromatic-aromatic interaction of Tyr2-Trp9. This is consistent with the values obtained

here (approximately -5 kcal/mol) for the initial interaction energy of Tyr2-Trp9 as shown in Fig. 38. For more than 16 Å, the most stabilizing residue pairs within chignolin are Asp3-Thr6, Asp3-Gly7, Tyr2-Trp9, and Asp3-Thr8. Of those, the most stabilizing interaction is the Asp3-Thr6 pair. It shows stability for over half of the stretch with an initial energy of approximately -25 kcal/mol. The interaction becomes negligible at an end-to-end distance of $r_{ee} \approx 18$ Å. At that point, the Glu5-Gly7 interaction becomes that most stabilizing and maintains contact for the remainder of the pull.

4.3.2 Trpzip mutations

4.3.2.1 Energetics

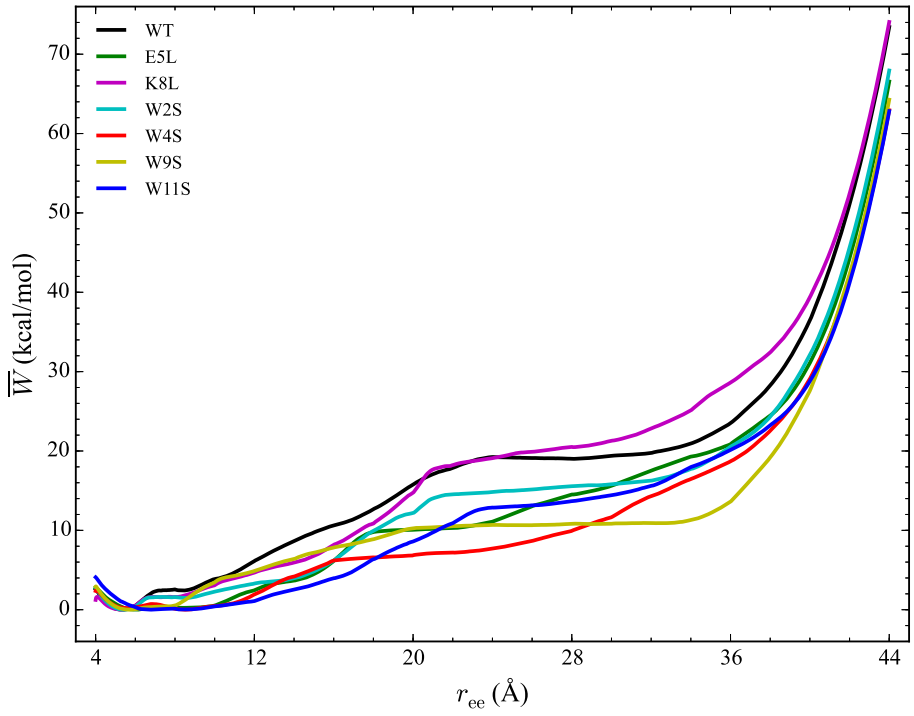

Figure 40: PMFs for each peptide in explicit solvent. The WT PMF is shown in black. The mutant PMFs are shown for comparison. Each PMF is obtained at a velocity of 1 Å/ns with a sampling size of 100 trajectories per stage.

Table 5: The ΔG values at structurally significant distances of r_{ee} (denoted across the top) for each peptide (on left side) based on the PMF curves from Fig. 40.

r_{ee} (Å)	4	12	20	28	36	44
WT	2.55	6.12	15.73	19.02	23.50	73.43
W2S	2.81	3.29	12.17	15.57	20.54	67.94
W4S	2.37	1.93	6.84	9.93	18.70	62.94
W9S	2.91	4.86	10.28	10.85	13.60	64.27
W11S	4.09	1.09	8.60	13.68	20.13	62.84
E5L	2.71	2.44	10.07	14.52	20.84	66.56
K8L	1.23	4.65	14.71	20.52	28.61	74.06

The work required to unfold the peptide varies along the unfolding pathway. This variation is expressed as the PMF for a specified reaction coordinate, namely the end-to-end distance r_{ee}. The energetics of the mechanical unfolding of the WT trpzip1 and six mutants in explicit solvent are shown in Fig. 40. The PMFs were obtained using the ASMD method with 100 trajectories per stage, at a pull velocity of 1 Å/ns. We have shown previously [31], using a pulling velocity of 1 Å/ns converges and is reliable for determining the overall trend in energetics between peptides. In Fig. 40, the PMF of the WT peptide is shown in black. The mutations, shown in several colors described in the legend of Fig. 40, substantially affect the magnitude and overall curvature of the PMFs indicating a difference in stability due to mutation. The mutations exhibit both stabilizing and destabilizing trends in the energetics. In particular, the K8L mutant (magenta curve), is the only mutation which results in a PMF that closely resembles the PMF of the WT peptide. For the first 20 Å of the pull, the WT peptide is the most stable until K8L surpasses the WT to become the most stable. This indicates that K8L maintains stability similar to that of the WT or even exceeding that of the WT. One comparison that can be made is between the salt bridge mutants K8L and E5L. In comparison to K8L, E5L is substantially destabilizing. It was originally hypothesized that both of those mutations would result in decreased stability considering that the mutations directly effect the salt bridge formation between residues Glu5 and Lys8.

In addition to those trends, other general observations about the energetics of the mutations can be made from Fig. 40. All peptides, except W11S (blue), have a similar minimum

at $\approx$ 5 Å. The W11S mutant has a very broad, not well-defined minimum. In the WT peptide, at $\approx$ 24 Å, the PMF peaks and begins to plateau. Interestingly, in every mutant, except for W11S, this "peak" is shifted to the left. Also, for the first half of the pull, the W11S mutant maintains the lowest PMF. Two of the mutants, E5L and W2S, show signs of a second "minimum" at $\approx$ 14 Å. Overall, it is clear that the mutations induced different energetic behaviors of the peptides during mechanical stretching simulations. These differences have been quantified and are comparable to experiments.

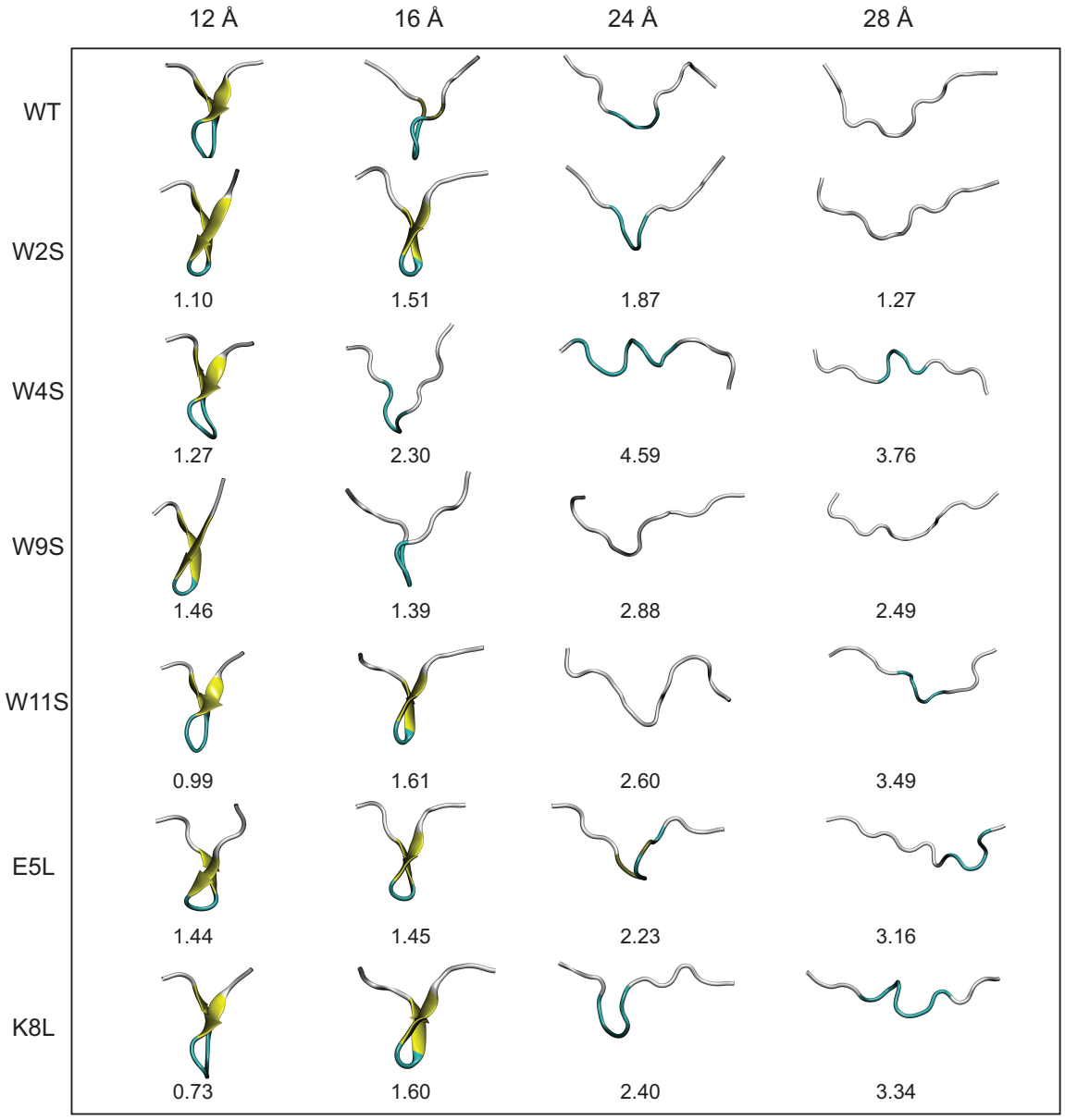

Figure 41: Illustrations of the unfolding pathways of each peptide. The peptides are colored by their representative secondary structures motifs. The coloring scheme is the following: Yellow represents β-hairpin structure, cyan corresponds to turn structure, and white is random coil structure. The RMSD values obtained, in Å, are listed below each structure. The RMSDs are calculated using the WT structure and a mutant structure at the corresponding stretching distance (top).

One useful metric to evaluate the trend in the free energies obtained for each of the peptides, is to measure and compare the values from the PMF at several different distances of r_{ee} as shown in Table 5. These trends are comparable to the experimental trends obtained in Table 6, except that the experimental trends do not include E5L. At a r_{ee} of 4 Å the order of ΔG values is W11S >W9S >W2S >E5L >WT >W4S >K8L. This trend shows that in the beginning of the pull, the Trp to Ser mutants (with the exception of W4S) are all stabilizing. At a r_{ee} of 12 Å the trend of the ΔG values inverses and the Trp to Ser mutants are less stabilizing: WT >W9S >K8L >W2S >E5L >W4S >W11S. It should also be noted that the WT peptide becomes the most stable peptide. At a r_{ee} of 20 Å the WT remains the most stabilized peptide with K8L as a close second: WT >K8L >W2S >W9S >E5L >W11S >W4S. At 28 Å there is a turn over between WT and K8L with K8L becoming the most stabilized peptide, which is consistent for for the remainder of the pull: K8L >WT >W2S >E5L >W11S >W9S >W4S. Overall, the last three sets of trends remain very similar. The only differences between the trends at 28 Å and 36 Å is that the E5L and W2S switch places and the W4S and W9S also switch places: K8L >WT >E5L >W2S >W11S >W4S >W9S. At the end of the reaction coordinate, at 44 Å, the order of the peptides is: K8L >WT >W2S >E5L >W9S >W4S >W11S. Again, K8L and WT remain the most stable peptides while W9S, W4S and W11S are the least stable. The order of the stability of the first four peptides in the same as the order at 28 Å. The only difference between the two trends, at 28 Å and at 44 Å, is that the W9S and W11S are reversed.

Another useful metric for discussing the effect of the mutations on the overall structural changes and pathways the peptides undergo during forced unfolding, is the calculation of Root Mean Square Deviations (RMSDs). The RMSDs provide a comparative tool to determine how different the secondary structure is between the WT and the mutant peptide. Snapshots of each peptide were taken along the trajectory at specific values of r_{ee} (i.e. 12, 16, 24, 28 Å). These values of r_{ee} correspond to features observed in the PMFs in Fig. 40. The snapshots of the structures and the corresponding RMSD obtained for that structure versus the WT are shown in Fig 41. Overall, all of the peptides maintain the β-hairpin structure for the first 8 Å of the pull. The r_{ee} with the narrowest distribution of RMSD

values sampled occurs at 12 Å. At that r_{ee}, K8L has the smallest RMSD value, 0.73 Å, which indicates that the mutant structure is most similar to the WT structure. This reinforces findings that the K8L mutant is most similarly stabilized to the WT peptide. At 12 Å, the mutants with the largest deviations (i.e. least similar to the WT) are W9S and E5L. It can also be observed the the WT, W4S and W9S mutants begin to lose their β-hairpin structure at 16 Å. The broadest range of RMSD values sampled occurs at the next r_{ee} of 24 Å. Notably, the largest deviation from the WT structure at 24 Å occurs with the W4S mutant, which has an RMSD value of 4.59 Å. It should also be noted that at 24 Å, no peptide exhibits any residual β-hairpin structure. At a r_{ee} of 28 Å the RMSDs cover also cover broad range from 1.27 (W2S) to 3.76 Å (W4S). At this distance, which is more than halfway through the reaction coordinate, several of the mutations have turn character. Specifically, the W4S, W11S, E5L and K8L structures are still being stabilized in specific regions.

4.3.2.2 Hydrogen bonds

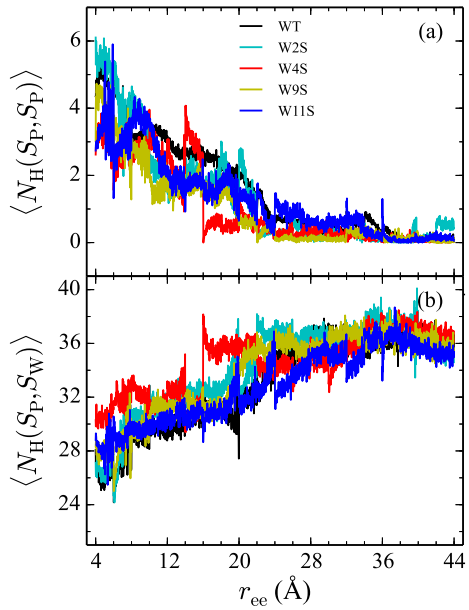

Figure 42: The weighted average number of intrapeptide hydrogen bonds for trpzip1 WT and the Trp to Ser mutants are shown. The weighted average number of hydrogen bonds formed within the peptide are shown in panel (a) and the bonds between the peptides and the explicit solvent is shown in panel (b). For each panel, the average is of 100 trajectories per stage at pulling velocity of 1 Å/ns.

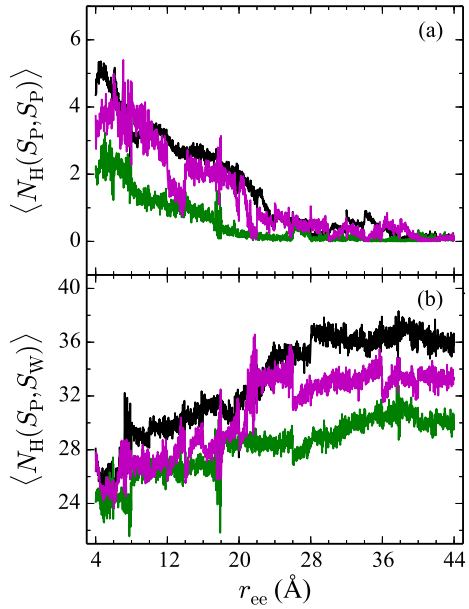

Figure 43: The weighted average number of intrapeptide hydrogen bonds for trpzip1 WT and the salt bridge mutants are shown. The weighted average number of hydrogen bonds formed within the peptide are shown in panel (a) and the bonds between peptides and the explicit solvent is shown in panel (b). For each panel, the average is of 100 trajectories per stage at pulling velocity of 1 Å/ns.

Hydrogen bonds have been shown to play a significant role in the determination of pathway and structure of peptides [189]. As shown in Fig. 42, the mutation of the tryptophans to serine not substantially effect the hydrogen bonding patterns in explicit solvent. However, the salt bridge mutations E5L and K8L do affect the both the hydrogen bonding patterns within the peptide and between the peptide and explicit solvent as shown in Fig. 43. In panel (a), the initial mutant structures have fewer bonds than the WT peptide. The WT begins with 5 intrapeptide hydrogen bonds, while K8L begins with 3 and E5L begins with 2. At a r_{ee} of 28 Å, the mutants and the WT intrapeptide bonds taper off to 0 bonds. The largest difference in the patterns is shown in panel (b). Though the peptides begin with about 24 bonds to solvent, by the end of the unfolding pathway the E5L mutant (green) forms 6 fewer bonds to solvent than the WT peptide (black). This indicates that the peptide

along the pulling pathway is not being stabilized by solvent during the unfolding transition and could be one of the factors contributing to the decrease in stability of the E5L mutant. On the other hand, K8L forms only 3 fewer bonds to solvent than WT, leading to a much more similar unfolding pathway and stabilization to the solvent that is experienced by the WT.

The Trp-to-Ser mutants all begin with ≈26-31 hydrogen bonds to solvent and end with ≈36 contacts to water as shown in Fig. 42c. This range is, on average, more the the WT and salt bridge mutants. At a r_{ee} of 20 Å, the K8L and E5L contacts to solvent begin to differentiate. By the end of the pull, the WT peptide has 36 bonds to solvent, K8L has 33, and E5L has 30. Overall, the hydrogen bond patterns are not as sensitive to the serine mutants, but they are sensitive to the mutations of the salt bridging residues.

4.3.2.3 Residue pair interactions

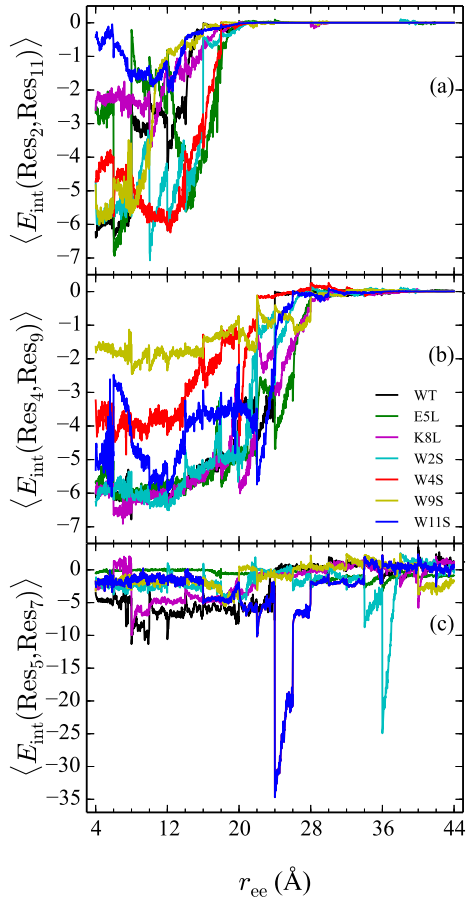

Figure 44: Comparison of the 2-11 (a), 4-9 (b), and 5-7 (c) interactions of trpzip1 mutants. The energies for each curve are obtained using a weighted average of 100 trajectories per stage at a velocity of 1 Å/ns.

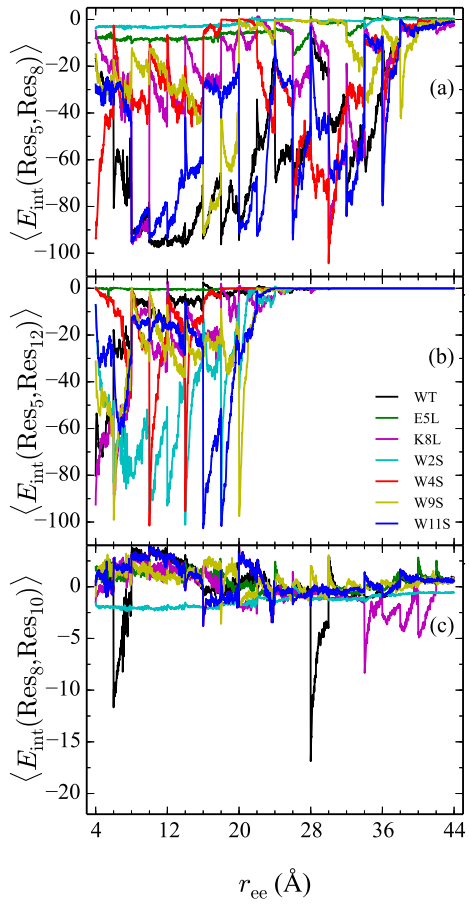

Figure 45: Comparison the 5-8 (a), 5-12 (b), and 8-10 (c) interactions of trpzip1 mutants. The energies for each curve are obtained using a weighted average of 100 trajectories per stage at a velocity of 1 Å/ns.

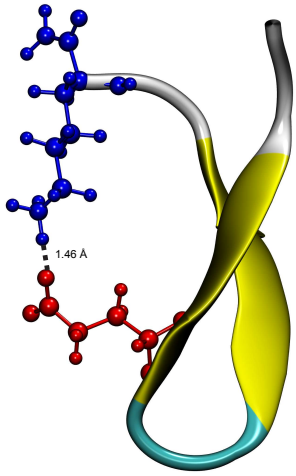

Figure 46: Unweighted average structure of K8L depicting the stabilizing interaction between Glu5 (red) and Lys12 (blue). The distance between the residues is shown as a dashed black line and was measured to be 1.46 Å. The overall r_{ee} is $\approx$ 8.6 Å

From the calculation of the persistence of specific residue pairs within each peptide, we can give an estimation of the order in which the contacts become non-interacting. All WT data presented in Figs. 44 and 45 have been reproduced from our previous work [31] comparing the WT trpzip1 with chignolin. In each figure, the WT interaction energy curves are shown in black. In the WT peptide, the most stabilizing interaction is the Glu5-Lys8 pair interaction seen in Fig. 45a. This finding was a result from our previous work [31] and has been reproduced for the current investigation. However, from the residue pair interaction energies of the salt bridge mutants, E5L and K8L, it can be shown that there is a competing salt bridge within K8L. As shown in Fig. 45b (magenta curve), the most stabilizing interaction within the mutant peptide K8L is Glu5-Lys12. The interaction is the most stable at the beginning of the pull in comparison to the WT (black curve), which lacks stabilization at that contact as shown in the illustration in Fig. 46. In this figure, K8L is shown with the peptide backbone and only the Glu5 (red) and Lys12 (blue) highlighted. The distance between the two residues, 1.46 Å, is shown as a dashed black line. The overall end-to-end distance of the peptide is 8.6 Å. This interaction occurs within K8L because the Glu5-Lys8 interaction is no longer available due to the mutation of Lys8 to Leu8. In

this interaction, the terminal lysine residue can interact with the glutamic acid, helping to stabilize the peptide during the stretch. This interaction is not possible in the E5L mutant because the negatively charged Glu has been mutated to a neutral Leu. In the WT peptide, the interaction is present but is not dominating because the Lys8 is positioned closer to Glu5. Due to the stabilizing nature of the competing salt bridge within the K8L mutant, the pathway of unfolding mimics that of the WT peptide. The Glu5-Lys12 is stabilized at the beginning of the pull at ≈ -80 kcal/mol and tapers off to 0 kcal/mol at a r_{ee} of 24 Å. The Glu5-Lys12 interaction is also an example of a nonnative contact formed during the course of unfolding that adds work required to stretch the peptide. It has been previously shown [207, 208] that the position of the charged residues within the peptide controls the stable structure adopted by the peptide; they can be favorable (resulting in a salt bridge) or it can be detrimental (diminishing hairpin formation). In a previous mutation study [218] of a similar secondary structure, protein GB1, it was found that cross-strand stabilization by side chain-side chain interactions were not the most dominant factor contributing to folded structures.

The interaction between residues at the 2 and 11 positions is shown in Fig. 44a. In the WT, the interaction is one of the two Trp-Trp interactions. Though the peptides range in stability from -1 to -6 kcal/mol, which generally indicates slight stabilization, it is apparent that the mutations lead to discernible trends in the stabilization. For example, the W11S mutant (blue curve) shows the most dramatic change in 2-11 interaction energy as was expected since the mutation of the tryptophan to serine at the 11 position directly effects that pair. The mutation is effectively cancelled out due to the mutation of the Trp11 to Ser. For each peptide, the interactions taper off at ≈ 20 Å.

This is also the case for the interaction energies between residues at the 4 and 9 positions, shown in Fig. 44b. In the WT, this interaction is also one of the two Trp-Trp interactions. The W4S (red curve) and W9S (yellow curve) mutants are more diminished for the 4-9 interaction as they are the peptides which were mutated to directly effect that residue pair interaction. All of the other peptides exhibit stabilization at ≈ -6 kcal/mol. As expected the interactions are maintained for a longer period of time than the Trp2-Trp11 interactions

and taper off at $\approx$ 28 Å.

The interactions between residues at the 8 and 10 positions shown in Fig. 45c, do not significantly contribute to the stability of the peptide. This is to be expected was to be expected because in the WT, the interaction is unfavorable between the Lys8 and Thr10. However, the WT peptide does so slight stability with the Lys8-Thr10 pair at r_{ee} between 6-10 Åand 28-32 Å. Interestingly, the K8L mutant also shows slight stabilization near the end of the pull at r_{ee} between 34-44 Å between the Leu8-Thr10 residue pair. The other profiles from the other mutant peptides does not show any stabilization.

Similarly, the interaction between residues 5 and 7 as shown in Fig. 44c only slightly contributes to stability of the peptide for all mutants. There is a slight spike in the stability of -35 kcal/mol for the Glu5-Asn7 in the W11S mutant at r_{ee} between 24-28 Å.

4.3.2.4 Experimental results and discussion

The far UV CD spectra of the peptides are shown in Fig. 47. The spectrum is characterized by a positive CD peak at approximately 228 nm that decreases in magnitude as a function of increasing temperature. Simultaneously, there is an increase in the magnitude of the negative CD signal at 213 nm. The thermal folding-unfolding reaction is reversible (data not shown) and cooperative, the spectrum is characterized by a high signal to noise ratio, and the thermally induced transition can be effectively monitored exclusively at 228 nm exiton-coupled band or by performing a global analysis using data between 200 nm and 250 nm. The near UV spectra of the peptides are shown in Fig. 48. Each spectrum is characterized by two predominate negative peaks at 285 and 295 nm. Both of these minima disappear with increasing temperature. The loss of signal at 295 nm and the global region between 260 and 315 nm are used to derive the near UV monitored thermodynamic parameters of thermal peptide unfolding.

The thermal behavior and far UV CD profile of the peptides is similar to data presented on other tryptophan zipper peptides by Cochran et al. [166] and to other trpzip mutants [219]. Peptide unfolding transitions are independent of peptide concentration and the forward and backward unfolding curves are superimposable (data not shown). Thermal

unfolding of the peptides fits well to a simple two state model (N $\leftrightarrow$ U) via analysis of the 228 nm CD signal as is shown in Fig. 49A. The exception to two state unfolding is the peptide E5L, which is best fit to a thermal unfolding model with a single intermediate: N $\leftrightarrow$ I $\leftrightarrow$ D. The thermal transition temperatures for the two state peptides, as well as other thermodynamic values, are shown in Table 6. For the far UV CD measurements, the order of stability is WT >K8L >W4S >(W9S = W11S) >W2S. Tm values range from 58.6°C (WT) to 33.4°C (W2S). This loss of stability is reflected in $\Delta(\Delta G)$ values that range from -0.55 to -1.16 kcal/mol relative to the WT peptide. Unfolding for all peptides is enthalpically favored at ambient temperatures (Table 6).

Near UV CD analysis monitors an earlier event in the unfolding of all trpzip-based peptides. Thermal unfolding of the peptides (except E5L) fits well to a simple two state model (N $\leftrightarrow$ U) via analysis of the 295 nm CD signal as is shown in Fig. 49B. Transition midpoint differences (relative to the T_m determined by analysis of the far UV CD data) range from 6.7°C for the WT peptide to 17.9°C for the W4S mutant. The rest of the mutants have a ΔT_m of approximately 13 to 14°C (Table 6). Thermal near UV CD is commonly used to monitor the loss of tertiary and side chain packing interactions versus the loss of secondary structure that is monitored by thermal far UV [220, 221, 222]. The tertiary packing stability of the peptide core, as measured by the near UV CD $\Delta(\Delta G)$ values in Table 6, gives a stability order: WT >K8L >(W9S = W11S) $\approx$ (W2S = W4S).

The E5L peptide is unusual in that there is a marked intermediate in the unfolding curve in the range of 27 to 39°C (Figs. 47,49A) and a smaller intermediate from 34 - 39°C (Figs. 48,49B). The nature of this intermediate is unknown, but it does not appear to be an aggregate as changes to buffer conditions (e.g.- addition of salt) does not alter thermal unfolding behavior (data not shown). Because of the unknown nature of the intermediate only the T_m values are presented in Table 7.

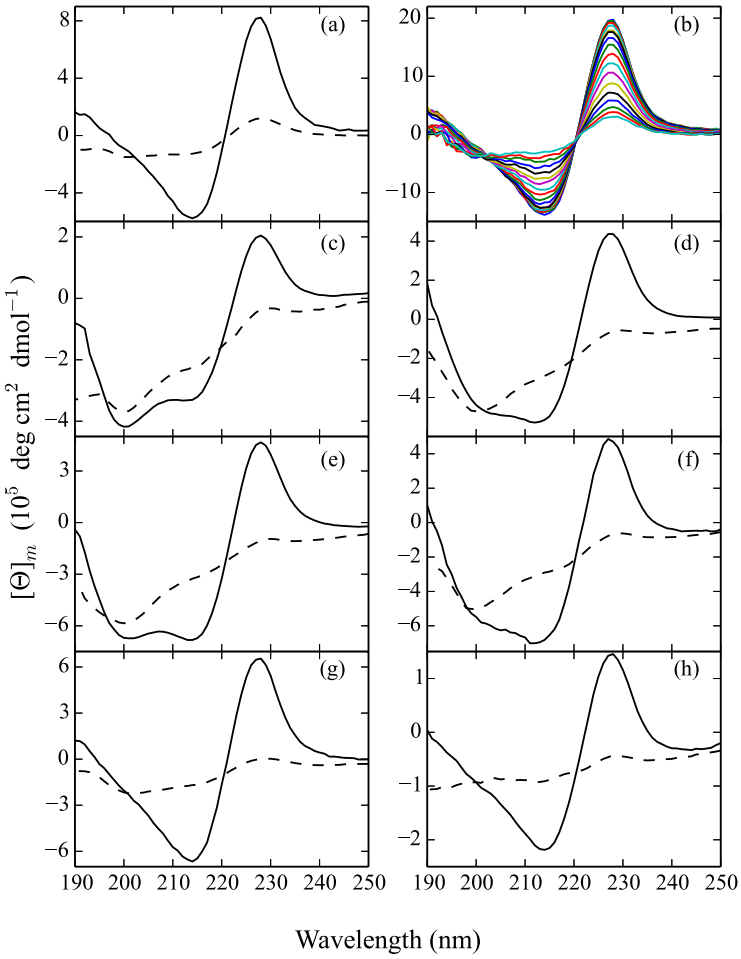

Figure 47: Far UV CD spectrum of the peptides at 5°C (solid black curves) and 90°C (dashed black curves). The panels correspond to the following peptides: A) Wild type, B) full far UV spectrum of the WT peptide as a function of temperature between 5 and 90°C plotted every 5°C, C) W2S, D) W4S, E) W9S, F) W11S, G) K8L, H) E5L. The temperature increase is marked by a loss of ellipticity at 228 nm and 212 nm. The global analysis to determine the peptide multi-wavelength melting temperature was performed between the 200 nm inflection point and 242 nm. This data was contributed by Dr. Stephen Quirk.

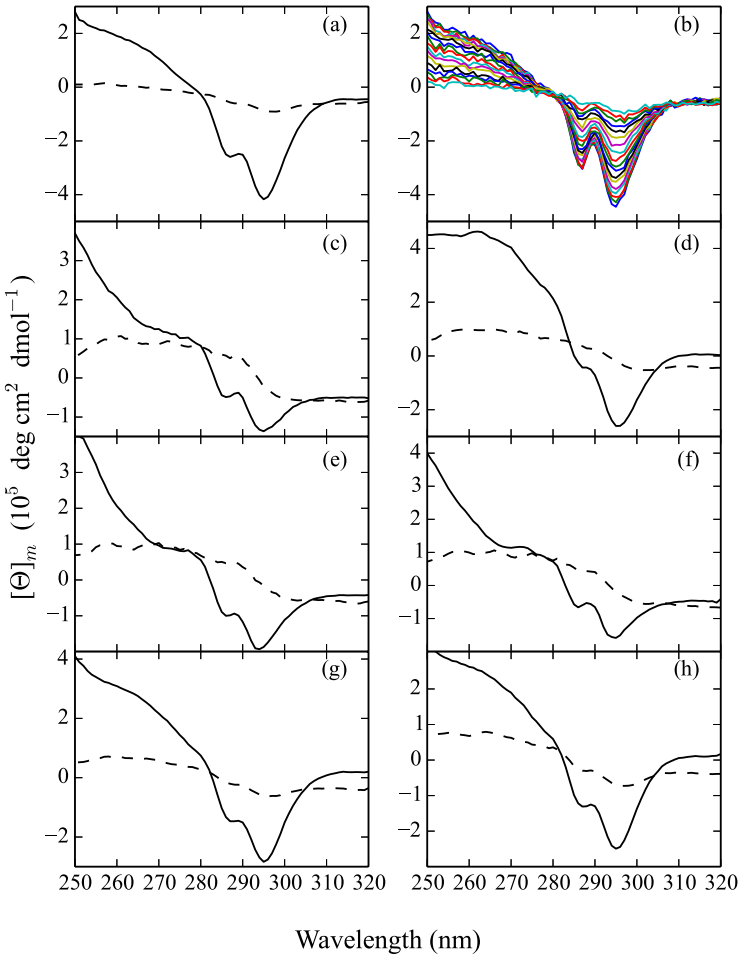

Figure 48: Near UV CD spectrum of the peptides at 5 °C (solid black curves) and 90 °C (dashed black curves). The panels correspond to the following peptides: A) WT, B) full near UV spectrum of the WT peptide as a function of temperature between 5 and 90 °C plotted every 5 °C, C) W2S, D) W4S, E) W9S, F) W11S, G) K8L, H) E5L. The temperature increase is marked by a loss of ellipticity at 285 nm and 295 nm. The global analysis to determine the peptide multi-wavelength melting temperature was performed between the 260 nm and 315 nm. This data was contributed by Dr. Stephen Quirk.

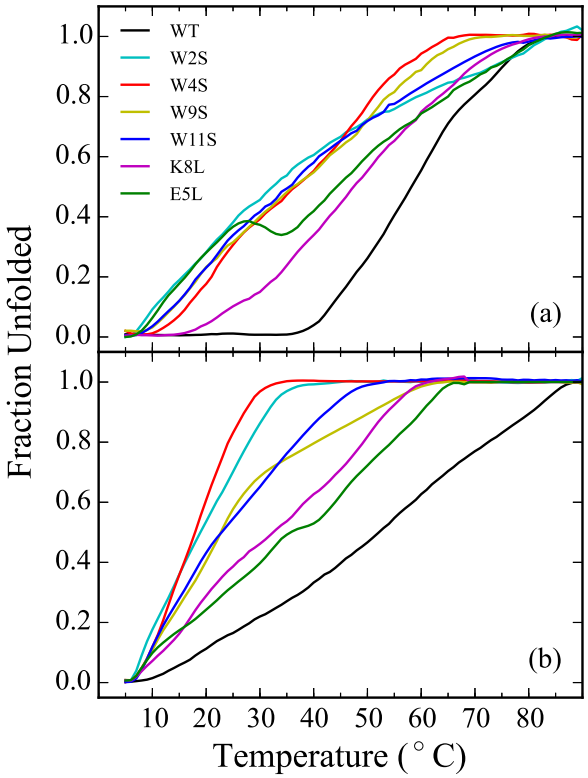

Figure 49: Fraction of unfolded peptide as a function of temperature calculated using data from the CD signal at 228 nm (a) or 295 nm (b) and Eq. 10. Line colors for both graphs: Wild type, black; W2S, cyan; W4S, red; W9S, yellow; W11S, blue; K8L, purple; E5L, green. This data was contributed by Dr. Stephen Quirk.

Table 6: Thermodynamic parameters for the thermal unfolding of peptides based on global analysis of the CD data in the far and near UV regions. The melting temperature is measured in °C, ΔH_m is in units of kcal mol^{-1}, ΔS_m is in units of cal mol^{-1} deg, and $\Delta(\Delta G)$ is in units of kcal mol^{-1}. This data was contributed by Dr. Stephen Quirk.

UV Region	Peptide	T_m	ΔH_m	ΔS_m	$\Delta(\Delta G)$
Far	WT	58.6 +/- 0.4	15.2 +/- 1.8	45.8 +/- 0.7	
	W2S	33.4 +/- 0.2	14.7 +/- 0.9	48.1 +/- 0.8	-1.16 +/- 0.07
	W4S	37.2 +/- 0.2	16.8 +/- 1.2	54.3 +/- 0.6	-0.99 +/- 0.06
	W9S	35.7 +/- 0.1	13.2 +/- 1.1	42.8 +/- 0.9	-1.06 +/- 0.08
	W11S	35.9 +/- 0.2	12.6 +/- 1.5	41.0 +/- 0.8	-1.05 +/- 0.08
	K8L	46.7 +/- 0.4	15.1 +/- 0.8	47.4 +/- 0.6	-0.55 +/- 0.04
Near	WT	51.9 +/- 0.3	13.7 +/- 1.2	42.4 +/- 0.5	
	W2S	19.3 +/- 0.1	11.2 +/- 1.0	38.4 +/- 0.6	-1.38 +/- 0.06
	W4S	19.3 +/- 0.1	14.3 +/- 1.3	49.2 +/- 0.6	-1.38 +/- 0.05
	W9S	22.6 +/- 0.2	11.4 +/- 1.0	38.7 +/- 0.7	-1.24 +/- 0.06
	W11S	22.8 +/- 0.2	10.9 +/- 1.3	36.9 +/- 0.6	-1.23 +/- 0.08
	K8L	32.8 +/- 0.3	12.2 +/- 1.4	40.1 +/- 0.7	-0.81 +/- 0.04

Table 7: Thermodynamic parameters for the thermal unfolding of the E5L peptide based on global analysis of the CD data in the far and near UV regions. This data was contributed by Dr. Stephen Quirk.

UV Region	Transition	T_m (°C)
Far	N $\rightarrow$ I	20.1 +/- 0.3
	I $\rightarrow$ D	57.5 +/- 0.4
Near	N $\rightarrow$ I	17.8 +/- 0.4
	I $\rightarrow$ D	53.8 +/- 0.2

4.4 Comparing the performance of CHARMM36 and Amber16 using the β-hairpin motif

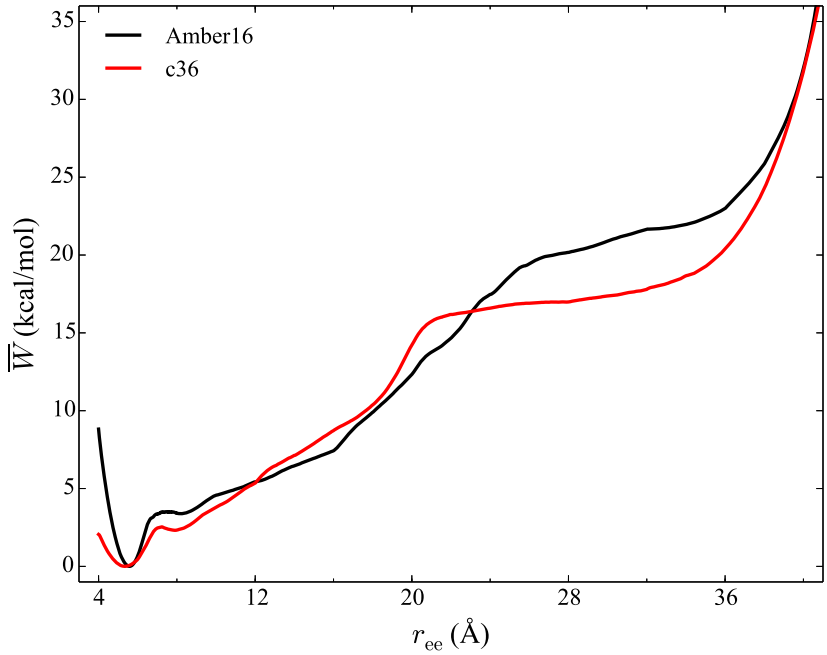

Figure 50: Energetic comparison of 1LE0 in both c36 and Amber16. The PMFs are obtained using ASMD with 100 tps at a pulling velocity of 1 Å/ns in explicit water solvent.

A comparison of the PMFs of trpzip1 calculated using c36 and Amber16 is shown in Fig. 50. There are several similarities and differences between the PMFs. The metric used to determine whether or not a difference is "significant" is the range of thermal fluctuation, generally 2-3 kT, covered by the PMFs at a given value of r_{ee}. First, the overall curvature and magnitude of the PMFs is similar between the two potentials. Both PMFs reach the same minimum structure at a r_{ee} of $\approx$ 5 Å and reach the same maximum at 35 kcal/mol by the end of the mechanical unfolding. The PMFs are also strikingly quantitatively similar in the regions of 4 to 20 Å and 36 to 40 Å. The first significant differences arises at the beginning of the pull. The Amber 16 PMF has a much more stabilized initial structure, by $\approx$ 9 kcal/mol, than the c36 result. The c36 PMF has a pronounced plateau beginning at 20 Å of extension. In the Amber16 PMF the plateau is still present, but it is shifted from

20 to ≈ 26 Å and it is not a well-defined as in the c36 PMF. It is also in this region that the the Amber16 PMF exhibits more stabilization. The c36 PMF plateaus at 17 kcal/mol for several angstroms of extension while the Amber16 PMF rises from ≈ 20 to 25 kcal/mol in this region.

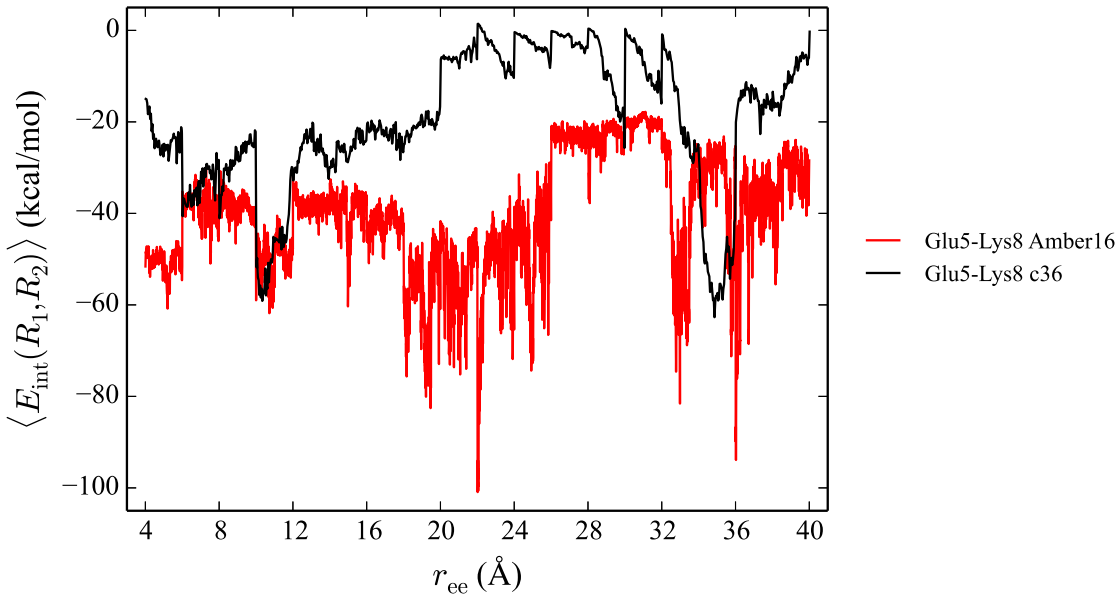

Figure 51: Comparison of the salt bridge formation within trpzip1 in both c36 (black curve) and Amber16 (red curve). The curves are obtained using weighted averages of ASMD simulations at 1 Å/ns using 100 trajectories per stage in explicit water solvent.

The comparison of the salt bridge interaction within trpzip1, Glu5-Lys8, calculated using the two different potentials is shown in Fig. 51. The interaction energies have been calculated using the weighted averaging scheme described in Chapter 2. The calculations are the average of 100 trajectories per stage using a pulling velocity of 1 Å/ns. There are several differences and similarities between the FFs characterization of the interaction. The Amber16 result (red curve), in general, shows a broader average than the c36 results (black curve). The first major difference is the energy of the initial interaction at a $r_{\rm ee}$ of 4 Å. The interaction in Amber16 is stabilized by ≈ 25 kcal/mol when compared to the c36 result. In the next region of the pull, from 6 to 12 Å, the two potentials show good agreement in

terms of similar values of stabilization. However, toward the middle region of the reaction coordinate, from ≈ 12 to 30 Å, the two curves diverge. In Amber16, the interaction is much more stabilized from a range of -40 to -80 kcal/mol. The interaction in c36 oscillates between 0 and -20 kcal/mol within this region. This discrepancy between the interaction energies could lead to one strong reason why the PMFs also deviate between the within the region with the Amber16 PMF being more stabilized than the c36 PMF as shown in Fig. 50. At the r_{ee} of 30 Å, the two curves become more similar. The end of the pull mirrors the beginning of the pull in that the c36 result is less stabilized, at ≈ 0 kcal/mol, than the Amber16, at ≈ 35 kcal/mol. Overall, this analysis indicates that there are several differences in calculated interaction energies for at least one interaction within trpzip1.

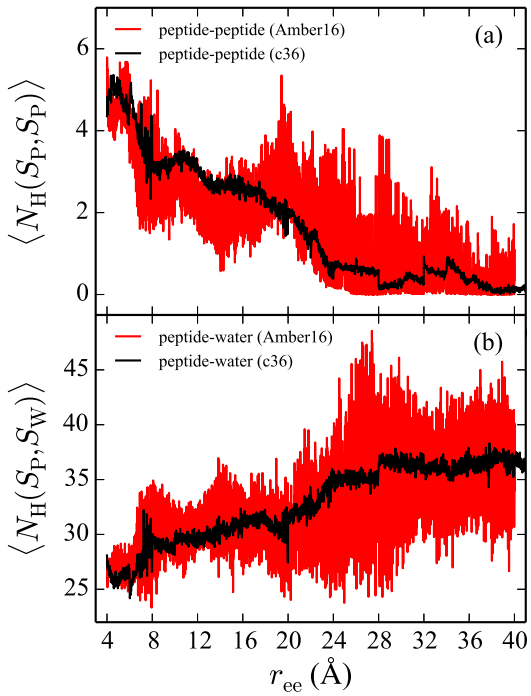

Figure 52: Comparison of the hydrogen bond formation and dissolution within trpzip1 in both Amber16 (red curve) and c36 (black curve). The intrapeptide hydrogen bonds are shown in the top panel and the peptide-water hydrogen bonds are shown in the bottom panel. The curves are obtained using weighted averages of ASMD simulations at 1 Å/ns using 100 trajectories per stage in explicit water solvent.

A comparison of the hydrogen bond patterns of trpzip1 obtained using Amber16 and c36 over the course of the forced unfolding reaction coordinate can be seen in Fig. 52. Similar to the trends observed with the salt bridge interaction in Fig. 51, the hydrogen bond patterns show several similarities and differences between the performances of the potentials. In general, the overall trends observed, such as the decrease in the number of intrapeptide hydrogen bonds from 5 to 0 and the increase is peptide water bonds from 26 to $\approx$ 35, remain the same. The largest difference is that the Amber16 results contain a broader average of hydrogen bonds, particularly in the peptide-water patterns shown in red in Fig. 52b. The range of the hydrogen bond patterns using c36 are much narrower.

These hydrogen bond patterns aid in the explanation of the plateau region of the PMFs shown in Fig. 50. The number of intrapeptide hydrogen bonds decreases from 2 to 0 near an extension of 20 Å. This coincides with the r_{ee} at which the plateau region begins in the PMFs.

4.5 Conclusions

4.5.1 Trpzip1 versus chignolin

In this work, we have captured and compared the relative energetics of two small, stable β-hairpin structures using Adaptive Steered Molecular Dynamics (ASMD). Although the two peptides under consideration, trpzip1 and chignolin, have very similar secondary structures and hydrogen bonding profiles, their stability and melting temperatures are quite different. The computed force profiles and observables shed light on the differences in stability over the course of the forced stretching. The forced unfolding energetics of trpzip1 are more stable than chignolin, requiring more work to stretch the peptide from an initially folded state to an unstructured state. However, it is much more difficult to converge the PMF of trpzip1 in comparison to chignolin, even at low stretching velocity. Convergence of chignolin was obtained with 10 trajectories per stage at 0.1 Å/ns, while trpzip1 required 100 trajectories per stage for convergence at that velocity. This may be due to stabilizing interactions, such as more hydrogen bonds to solvent or salt bridging effects within trpzip1 that are not found in chignolin.

The solvent environment plays a role in the convergence and shapes of the energetics. The vacuum trpzip1 PMF converges at much higher stretching velocity than it does in explicit solvent. This trend was also observed in our previous work with α-helices [27, 55, 56], and by others [94, 223]. In all of these cases, the forced unfolding of helical peptides in vacuum was found to be less energetically favorable due to a lack of stabilizing contacts that are normally provided by water. Our structural results complement the findings of Schulz, Miettinen, and Netz [224] with respect to the dynamical response of the peptide. Specifically, they suggest that the frictional unfolding of α-helices is larger than that of β-hairpins because of varying degrees of the hydrogen bond strength. This is due to steric

effects from the solvent and amino acid side chain environment which is consistent with the structures that we found along our forced unfolding pathway.

In addition to energetics, we have obtained quantitative insight into the stability of our target peptides through an analysis of the intrapeptide and peptide-solvent hydrogen bond structure. At the end of the stretching of trpzip1 in vacuum, it contains approximately 4 hydrogen bonds but they are primarily not the originally hydrogen bonding pairs in the β-sheet structure. These bonds add to the overall stability of the peptide in vacuum but at the cost of steric entropy that can be relaxed in water where more favorable peptide-solvent hydrogen bonds can be formed instead. This leads to a lowering of the free energy to stretch the peptide in solvent. The same trend in hydrogen bonds is observed in chignolin as would be expected because the secondary structures are very similar. Similar effects due to solvent stabilization were also observed in earlier work on α-helices [55, 56]. In that work it was speculated that all-atom representations (explicit) of the water solvent was necessary so as to be able to accommodate the changes in the hydrogen bonding structure. However, we recently found that implicit solvents can also accommodate the structure formation due to the fact that they include local response to the exposed charges on the peptide [27].

The origins of the stability of the β peptides studied here can be understood through the specific residue-pair interactions along the stretching pathway. In trpzip1, the most energetically favorable residue interaction occurs between Glu5 and Lys8. In vacuum, this interaction is stable, even when the peptide is fully stretched. In explicit solvent, it remains the largest contributor to the stability of the fold, but the interaction is dampened due to solvent effects. However, the interaction is essentially 0 kcal/mol near the end of the unfolding path. In explicit solvent, the two tryptophan pairs, Trp2-Trp11 and Trp4-Trp9, also add stability to the peptide's hydrophobic core. Chignolin does not contain any salt bridges, but there are several residue pairs such as Asp3-Thr6 which provides stability for the first half of the stretch, and Glu5-Gly7 which add some stability toward the end of the unfolding. Other interactions, such as the Asp3-Gly7 and Tyr2-Trp9 also contribute to the overall unfolding pathway of chignolin. In contrast to previous work on helical peptides [94], which found pronounced stabilization of an alanine-rich helix as Glu-Lys salt bridges

were introduced to the $i \rightarrow i+3$ contacts positions, we found that the β-hairpin motif is significantly stabilized by the addition of salt bridges.

In conclusion, the lowering of the PMF, from vacuum to explicit solvent, of the unfolding of trpzip1 is related to solvent effects experienced by the Glu5-Lys8 electrostatic interaction and on the loss of intrapeptide hydrogen bonding contacts to water. Due to the stabilization of the Glu5-Lys8 interaction in vacuum, the PMF of trpzip1 is significantly higher than in explicit solvent. This favorable salt bridge is also, in large part, why the PMF of trpzip1 is considerably higher than that of chignolin. The turn region in trpzip1 is supported by the Glu5-Lys8 salt bridge. This could provide some explanation of why chignolin and trpzip1, although similar in structure, are not similar with regard to stability (as shown by melting temperature). Undoubtedly, the four Trp residues within trpzip1 also play a role [171] in creating a hydrophobic "core"; however, it remains debatable whether the stability of trpzip1 is driven by hydrophobic collapse, turn stability, or some combination of these two effects. This could be resolved through a mutation study involving the salt bridging residues and several other residues within chignolin. Another remaining question is concerned with how the length of the β-sheet strand affects the peptide stability [225].

4.5.2 Trpzip1 mutations

In this work, we have investigated six mutants of trpzip1 using experimental and computational methods to determine the stability of each peptide. We compare the computationally determined mechanical unfolding pathways of each mutant with the WT peptide. We have also calculated the side chain-side chain and hydrogen bonding effects within each mutant. From this analysis, we have found that the stability of competing salt bridge between Glu5-Lys8 and Glu5-Lys12 is within K8L is substantial and directly effects the energetics and unfolding pathway of the peptide. The mutations E5L and K8L effect the hydrogen bonding trends between intrapeptide and peptide-solvent bonds. There is evidence to suggest that there is an interplay between hydrophobic collapse and side chain-side chain interactions. The K8L mutant is the most stable because structurally and energetically it resembles the

WT peptide. This is also confirmed from the experimental CD spectra and thermal unfolding. Like the WT peptide, K8L is stabilized by a salt bridge. However, this contact is between the Glu5 and Lys12 residues, which is a nonnative contact in the WT peptide. In the case of E5L, this contact is not possible because the Glu at the 5 position was mutated to a Leucine residue.

4.5.3 Comparing the performance of CHARMM36 and Amber16 using the β-hairpin motif

The work highlighted in this study is complementary to the work in Chapter 3 focused on helical peptides. Here, we have created a direct comparison of the performance of the c36 and Amber16 FFs using the β-hairpin secondary structure motif. The results with β-hairpins reinforce our conclusions from Chapter 3 in that the use of the different potentials does not significantly effect the unfolding energetics and general secondary structure obtained for the peptides. Again, the overall magnitude and curvature of the PMFs are consistent between the two FFs; however, the depth of the initial minimum structure is much more stabilized in the Amber16 FF. This is considered substantial as it exceeds the normal thermal fluctuation range.

Numerous differences arise between the FFs in the evaluation of the interaction energy calculated between Glu5 and Lys8. This represents a significant test of the FFs as this is recognized as a stabilizing electrostatic interaction. The c36 result has a much narrower average than the Amber16 result. Near the beginning and end of the unfolding reaction coordinate, the salt bridge energy analysis is consistent between the FFs. However, in the middle region of the unfolding, the energies diverge substantially. The hydrogen bond patterns are another means of comparing the performance of the two FFs. The overall trends in the hydrogen bond patterns are consistent; however, the c36 result is much narrower than the Amber16.

This work has provided a direct comparison using β-hairpin peptides to assess the performance of the two most popular MD FFs, Amber16 and c36. While the FFs can be used reliably to capture the energetics of unfolding, caution should be used when investigating the specific interactions within the system that may effect the overall pathway of unfolding.

CHAPTER V

ENERGETICS OF UNFOLDING MULTI-MOTIF PEPTIDES

5.1 Introduction

5.1.1 Neuropeptide Y (NPY) mutations

NPY [226] is a multi-functional protein, consisting of 36 residues, found in abundance in the brain and nervous system. It is known that the peptide can bind to at least six different types of G-coupled protein receptors [227]. It has been implicated as a control switch for several biological processes such as regulation of food intake, vasoconstriction, memory retention and sleep regulation [228, 229]. As such, neuropeptides have been tagged as potential drug targets for treating diseases such as epilepsy, obesity, neurodegenerative and psychiatric disorders due to their bioavailability and ability to penetrate the blood-brain barrier [230, 231]. However, much is left unknown about how to modulate the stability the neuropeptides and how those changes in stability affect their function.

There have been several experimental and theoretical studies aimed at deconstructing the roles that each residue or secondary structure motif plays in the stability and function of the peptide. An early CD and NMR study by Nordmann et al. [232] determined that the stable form of NPY as a monomer was a polyproline helix. In contrast, a CD and FRET study by Bettio et al. [233] suggested that NPY as a monomer does not exist in a stable polyproline helical fold. However, a NMR study in which it was determined that the rigidity is not an intrinsic property of the peptide but instead it is due to the apt positioning of proline residues capable of stabilizing the polyproline fold [234]. The authors also comment on the role of Y20 and Y27 as stabilizing residues that help maintain the hairpin-like structure of the polyproline fold. This source of rigidity is further investigated by a molecular simulation using Langevin dynamics aimed at understanding the behavior of NPY under constant external force determined that NPY relaxes to an equilibrium structure at a fast timescale due to the rigidity of the polyproline helix motif[235]. One

recent mass spectrometry study indicated that the second proline from the N-terminus (found in a variety of neuropeptides) can undergo cis-trans isomerization, allowing the peptide to achieve a high level conformational heterogeneity [236]. In our previous work on the "unhinging" mechanism of NPY [54], which is integral for understanding the binding pathway of the protein to various receptors, was investigated at low and high temperatures. It was determined that the rate at which NPY unfolds at 310 K is too slow to occur naturally; therefore, NPY exists in a stable, folded confirmation at body temperature.

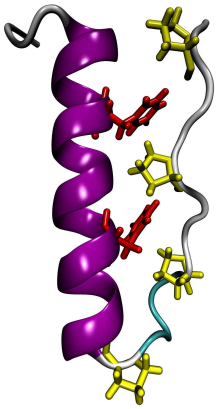

Figure 53: Illustration of the double mutant of NPY, Y2027F. The secondary structure of the peptide is depicted as a ribbon where the helix is in magenta, the polyproline turn is cyan and the unstructured tail is white. Several residue side chains are highlighted: phenylalanines (red) and proline (yellow).

In the present study, the stability of the WT NPY has been modulated using mutations of the tyrosine residues to phenylalanine at the 20 and 27 positions. An illustration of the double mutant peptide, Y2027F, is shown in Fig. 53 where the helix is in magenta, the polyproline turn is cyan, and the unstructured tail is white. The residues of interest, namely the phenylalanines 20 and 27, are highlighted in red while the prolines, which are known to play a significant role in the peptide structure are in yellow. Using a combination of molecular dynamics, thermal unfolding experiments, lipid and receptor binding kinetics studies, and analytical ultracentrifugation experiments, we are able to span length scale from the single molecule level to the quaternary structure in order to further understand

the cause of the increase in stability. On the single molecule scale, the opening of the hydrophobic pocket and hydrogen bonding contacts are also obtained and reinforce these findings. These mutations increase the thermodynamic stability (T_m) of the peptide by as much as 30°C. The determination of the energetics and pathways along the pulling coordinate will reveal the specific interactions stabilizing NPY peptides, particularly the mutants.

5.1.2 Enterocin 7B mutations

Enterocin 7B (PDB ID: 2M60) [237] is a small bacteriocin antimicrobial peptide composed of 43 amino acid residues. It was only in 2013 that the NMR solution structure of the peptide became available. Bacteriocins are peptides that are produced by bacteria that have antimicrobial properties [238]. A majority of bacteriocins are referred to as a "leader" peptide because they are produced with an N-terminal extension. These leader sequences have been implicated in a wide variety of processes involving transporting bacteriocins. Enterocin 7B is a "leaderless" bacteriocin, meaning it does not contain an N-terminal extension. However, leaderless bacteriocins have a broader range of activity than leader bacteriocins and are therefore of significant interest. It has been observed the that Enterocin 7B adopts a stable structure composed of three, short α-helices separated by turn regions [237]. Primary sequence analysis indicates that all three helices have varying degrees of amphipathic character. Interest in using Enterocin 7B as a food preservative has grown as it is effective at killing a large range of gram-positive bacteria including several human food-borne pathogens. Unlike most of the other enterocins, 7B is fully structured in solution. It is unknown how this peptide transverses through membranes in order to disrupt the cellular structure of bacteria; hence, resulting in cell death. We hypothesize that, similar to NPY, the peptide is stabilized primarily through a well-packed hydrophobic core. One major question that arises is how the stability of the hydrophobic core effects the overall dynamic behavior and function of Enterocin 7B.

In this work, which is still in progress [34], we approach this question from the computational perspective. ASMD simulations are currently underway using WT Enterocin 7B and

ten mutations. These simulations will probe the structural rigidity of Enterocin 7B. Details of the mutations and subsequent simulations are discussed in Sec. 5.2.2. Preliminary results of the energetics are presented in Sec. 5.3.2.

5.2 Peptide model details

The peptides investigated in this Chapter are WT NPY and three mutants along with WT Enterocin 7B and ten mutations. In the following sections, a detailed description of the design of the systems and simulation parameters used will be presented.

5.2.1 NPY mutations

5.2.1.1 Computational model

The solution NMR structure of the folded conformation of the WT NPY [226] was used as the structure from which the mutants were computationally prepared. Three mutants were built using the VMD plugin Mutator [63]. The specific mutations are denoted as Y20F, Y27F and Y2027F. Y20F and Y27F are both single mutations of the tyrosine at the 20 and 27 positions to phenylalanine. The double mutant, Y2027F, consists of two mutations of tyrosine to phenylalanine at the 20 and 27 positions. Phenylalanine was chosen as the substituting residue due to a need to increase the hydrophobicity of the pocket. For clarity, the sequences of the peptides used in this study are listed below:

1. WT: YPSKPDNPGEDAPAEDMARYYSALRHYINLITRQRY
2. Y20F: YPSKPDNPGEDAPAEDMARFYSALRHYINLITRQRY
3. Y27F: YPSKPDNPGEDAPAEDMARYYSALRHFINLITRQRY
4. Y2027F: YPSKPDNPGEDAPAEDMARFYSALRHFINLITRQRY

As observed with the VMD plugin Timeline, there were no significant structural changes to the backbone of the peptides due to the mutations. All of the peptides present as stable, polyproline helices. Timeline analysis of the secondary structure was performed after the equilibration of each peptide as described below.

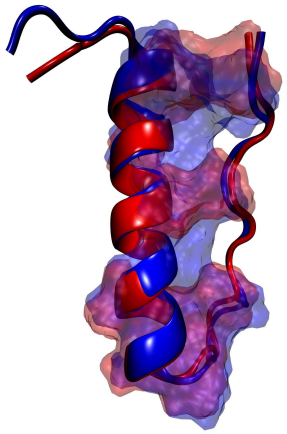

Figure 54: Illustration of WT NPY (red) overlaid with the Y2027F mutant (blue). The peptide backbones are represented as "ribbons". The hydrophobic residues of the peptides are represented as transparent surfaces.

Equilibration All of the ASMD simulations are performed using NAMD [139] as the molecular dynamics integrator with the c36 FF [96, 113]. All equilibrations and simulations therein described are performed in explicit water solvent using the TIP3P solvation algorithm as implemented in the NAMD solvation package. First, the peptides are solvated individually is a square cuboid box with approximately 19,000 water molecules. The solvent box is built with two equal sides of length L_{xy}, and a longer side along the z-axis of length L_z. The length of L_{xy} is 30 Å, and L_z=150 Å. After solvation, the overall charge of the peptide and solvent system is neutral. No ions are required to gain neutrality. After obtaining the solvated peptide, the system undergoes a three-step equilibration protocol. The first step of the protocol is to equilibrate the solvent and side chains of the peptide while keeping the backbone of the peptide restrained. This step allows the water to reach a density of approximately 0.9998 g/cm^3 while simultaneously allowing the side chains of the peptide to find favorable conformations. Energy minimization was performed for 10,000 steps to remove bad contacts using the conjugate gradient method. The system was equilibrated for 1 ns at 300K under NPT conditions. The Nose-Hoover Langevin piston method was used with a damping coefficient of 5 ps^{-1}, a decay period of 100 fs and a damping time constant

of 50 fs. Temperature is maintained using the Langevin thermostat. Other important simulation parameters are set as follows: the smooth switching function begins at 8 Å, van der Waals interaction cutoff distance was 12 Å, pair list distances are truncated at 13.5 Å, and long-range electrostatic forces were calculated using the particle-mesh Ewald summation method with a grid size of ¡ 1 Å. Those parameters are carried over to the second and third steps of equilibration, as well. In the second step of the equilibration protocol, the system undergoes a constant volume equilibration via the NVT ensemble. The backbone is initially restrained with a harmonic potential of 10 kcal/mol Å^2. The constraints on the backbone are gradually released during a series of 100 ps intervals. The constraints are gradually reduced to 5.0 and then 1.0 kcal/mol Å^2 consecutively. The final constraint of 1.0 kcal/mol Å^2 was then released and the entire system was allowed to equilibrate for another 500 ps. The gradual release of the constraints on the backbone of the peptide ensures that the peptide atoms do not begin to shift too quickly in reference to the solvent molecules.

Due to the free motion of the peptide during the second step of the equilibration protocol, the peptide can drift substantially from the z-axis (i.e. the axis along which the unfolding is steered). Therefore, in the third and final step of the equilibration the equilibrated peptide from the second step is removed from solvent and reoriented along the z-axis. This rotation does not in any way change the equilibrated structure, which was obtained in the previous step. Next, the rotated peptide is solvated using the same parameters as before. Following the solvation, the C_α atoms of the 1st and 36th residues are restrained but all other motions unconstrained. A 1 ns equilibration of the water and peptide system is then performed to equilibrate the solvent and side chains. This result was verified using the NAMD plug-in Timeline for analyzing secondary structure and determining root mean squared deviations.

Unfolding simulations In this work, the PMFs for the unfolding pathway of the four NPY peptides are obtained computationally using the ASMD as described in Chapter 2. This method was previously used to determine the mechanism by which WT NPY "unhinged" using a circular pulling pathway [54]. The efficiency of the algorithm has also been benchmarked for various solvent conditions including vacuum [55], implicit solvent [27], and

explicit solvent [56].

In all of the ASMD simulations, the peptides are stretched along the long z-axis to mimic unfolding. All simulations are carried out at a temperature of 300 K. The stretching coordinate is defined as the distance between the C_α of the 1st and 36th residues. At the start of each simulation, the distance between the stationary and pulled atom is constrained to 16 Å. This constraint is used to ensure that during the driven unfolding, the peptide accesses the local minimum obtained from the equilibration protocol. During the pull, the peptides are gradually stretched for 120 Å, resulting in a domain of the end-to-end distance from 16 Åto 136 Å. The ASMD simulations use forty evenly partitioned stages along the stretching coordinate. This results in each stage as a 3 Å pull. The PMF of the peptides is evaluated at a stretching velocity of 10 Å/ns with a sampling size of 100 trajectories per stage. Once the PMFs are obtained, they are shifted to reset the zero of energy to the minimum. This shifting, which allows for a more direct comparison of the PMFs, can be seen in Fig. 56.

The ASMD simulations require approximately 40 stages along the predefined reaction coordinate, with a sampling target of 100 trajectories per stage using steering velocities of 10 Å/ns and 1 Å/ns. There are a total of four peptides (WT and three mutants).

5.2.2 Enterocin 7B mutations

Beginning at the N-terminus, the first helix (denoted as Helix 1) is composed of residues 4 through 10. This is shown in blue in Fig. 55. The central helix, Helix 2, is comprised of residues 14-30 and is shown in red in the figure. However, Helix 2 deviates from regular helical ϕ-ψ space with a pronounced "kink" caused by residue Lys18. Helix 3 is comprised of residues 33 through 42 and is shown in yellow. Eleven different peptides are used for the simulations: WT Enterocin 7B and 10 mutants. A series of mutations were suggested by Dr. Stephen Quirk. Within Helix 1 three mutants were proposed as destabilizing mutants: I4R, L7R and V8R while one mutant, A5L, was proposed as stabilizing. In Helix 2, three mutants are proposed to be destabilizing, namely F15R, I23R and F26R while two mutations are suggested to be stabilizing, M24L and F26L. Similarly, in Helix 3 the L40R mutant was

proposed as destabilizing. For all of peptides, the initial r_{ee} is 20 Å after the three-step equilibration protocol (previously described in Sec. 5.2.1.1). Each peptide is individually solvated in a rectangular cuboid box using TIP3P explicitly represented water molecules with the minimal dimensions of 40Å x 40Å x 250Å. This is ≈ 11,500 water molecules for a total system size of ≈ 40,000 atoms. The ASMD simulations are performed using a reaction coordinate of 60 Å. This does not result in a completely unstructured peptides as the interest of the study is to investigate only the unstructuring of the hydrophobic core of the peptides. For the ASMD simulations 20 stages will be used to evenly descritize the reaction coordinate. This equates to 3 Å of unfolding within each stage. A sampling size of 100 trajectories per stage will be used along with a pull velocity of 1 Å/ns.

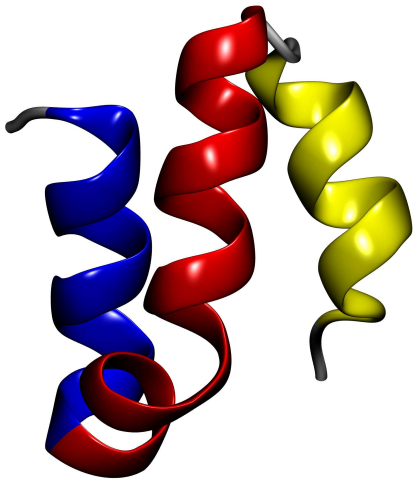

Figure 55: Illustration of WT Enterocin 7B. The secondary structure of the peptide is represented as a ribbons. The three different helices are highlighted in different colors. Helix 1 (residues 1-13) is blue, Helix 2 (residues 14-30) is red and Helix 3 (residues 33 -42) is yellow.

5.3 Results and Discussion

5.3.1 NPY mutations

Computational and experimental results of the energetics, structure, and stability of WT NPY and three mutants will be described in this section. All of the thermal unfolding

and receptor binding kinetics data presented in Sec. 5.3.1.4 and 5.3.1.5 was obtained by Dr. Stephen Quirk. The analytical ultracentrifugation data discussed in Sec. 5.3.1.6 was obtained by Dr. David Bain.

5.3.1.1 Energetics and convergence

The PMFs of the four peptides were obtained in order to compare and contrast the stability of each of the peptides. It is useful to approach the PMFs using references to specific regions of the reaction coordinate, r_{ee}, to compare the trends observed amongst the peptides. First, the addition of the mutations to the peptides creates a more stabilized local minimum, which can be seen in Fig. 56 at a r_{ee} of 20 Å. The Y20F (blue), Y27F (red), and Y2027F (green) PMFs all have more pronounced minima in comparison to the WT (black curve). The minimum of Y2027F is shifted to the right of the minima of Y20F and Y27F indicating that the local stable structure of the double mutant is more stable than either of the single mutants. In the region of the PMFs between 30 to 70 Å the clearest trend that emerges is the similarity between the Y20F and Y27F curves versus the similarity between the WT and Y2027F curves. Lastly, from 70 to 80 Å the PMFs begin to increase rapidly. This is the point at which the hydrophobic pocket is opening. At approximately 70 Å, the WT pocket begins opening first. The single mutant pockets open at 75 Å and finally the WT pocket opens at 80 Å. This trend is confirmed by a more detailed analysis of the interaction of the hydrophobic core interaction energy with the solvent detailed in Sec. 5.3.1.2.

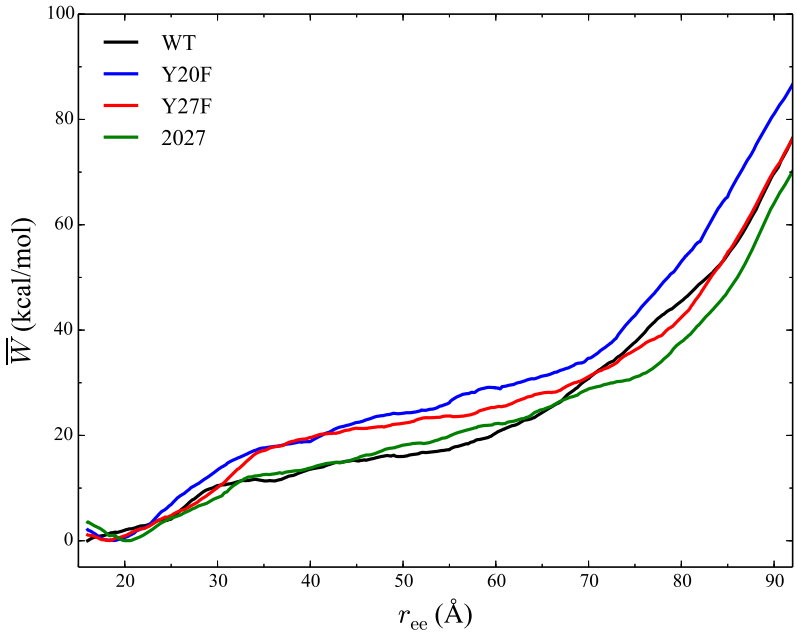

Figure 56: PMFs obtained for all for peptides using a sampling size of 100 trajectories per stage at a pulling velocity of 10 Å/ns in explicit water solvent. Only the first 70 Å of the pull are shown for clarity.

5.3.1.2 Opening of the hydrophobic core

The interaction energy between the hydrophobic residues within the peptides and the water can be seen in Fig. 57. The hydrophobic pocket is highlighted as a surface in Fig. 54. This result has been calculated using the weighted averaging scheme used for calculated the residue pair interaction energies and hydrogen bonding patterns described for observables in Chapter 2. The curves are the results of the weighted average of all the trajectories. Within the plots, information about the opening of the hydrophobic pocket can be obtained. The region between r_{ee} 60 and 80 Å is the range of end-to-end distances at which the hydrophobic pocket opens. In each case, after the pocket opens, the interaction of the pocket with the water increases substantially. The more stable the peptide, the more shifted this opening of the pocket. As shown by the black curve, the WT pocket opens first at a r_{ee} of 60 Å. The single mutants, Y20F (blue) and Y27F (green), have similar curves. They both have

a peak at approximately 75 Å. As shown in red, the double mutant, Y2027F, exhibits the most shifted opening of the pocket at a r_{ee} of 80 Å. All together these results create a clear trend for the opening of the hydrophobic pocket, and hence peptide stability, as follows: WT >Y20F/Y27F >Y2027F.

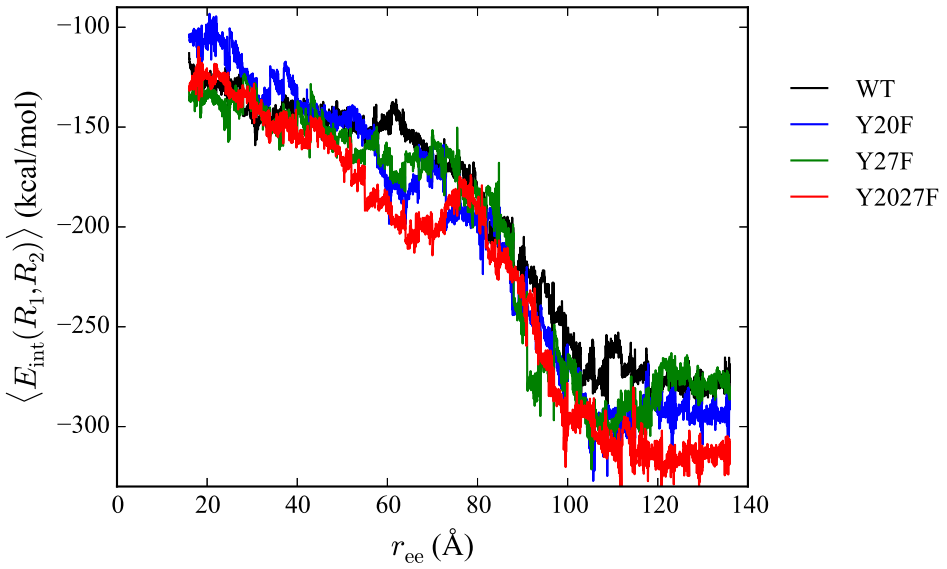

Figure 57: Comparison of the opening of the hydrophobic core across all peptides. The energies for each curve are obtained using a weighted average of 100 trajectories per stage at a velocity of 10 Å/ns in explicit water solvent.

5.3.1.3 Hydrogen bonding profiles

The weighted average of the different types of hydrogen bond patterns for WT NPY and the three mutants can be observed in Figs. 58, 59 and 60. Fig. 58 shows the average number of intrapeptide hydrogen bonds formed and broken over the course of the mechanical unfolding of the four peptides. Each of the peptides exhibits a similar trend in the number of hydrogen bond contacts. The initial structures all begin with 5 to 10 bonds. Between the region of 20 to 60 Å, the number of intrapeptide bonds formed increases to a range of 20 to 25 bonds. The WT, Y20F and Y27F peptides all have a maximum of ≈ 20 bonds at a r_{ee} of 70 Å. Y2027F forms the most contacts at a r_{ee} of 80 Å, after which the number of bonds formed begins to decrease rapidly until the end of the reaction coordinate. This coincides with

the behavior and trends of the opening of the hydrophobic pocket presented in Sec. 5.3.1.2. It should also be noted that the hydrogen bonding trend observed in Fig. 58 is consistent with $i \rightarrow i+4$ (α-helical) bonding patterns observed for the helical peptides investigated in Chapter 3.

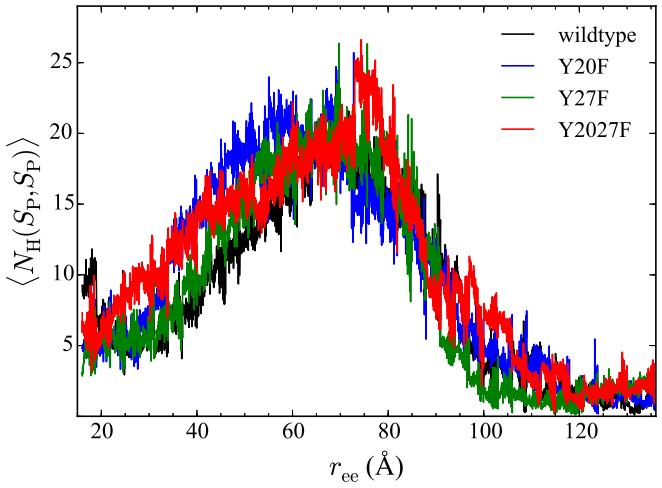

Figure 58: The weighted average intrapeptide hydrogen bonds of WT NPY (black), Y20F (blue), Y27F (green), and Y2027F (red). The average is of 100 trajectories per stage at pulling velocity of 10 Å/ns in explicit water solvent.

The hydrogen bonds formed between the peptide and the explicit water solvent are shown in Fig. 59. The WT and Y2027F mutant both begin with a slightly elevated number of hydrogen bond contacts with solvent at 85 contacts. The Y20F and Y27F mutants both begin with 80 contacts. For each peptide, the number of bonds formed to solvent remains constant over the first half of the unfolding. The number of bonds begins to increase at a r_{ee} of 80 Å and peaks at 110 bonds at a r_{ee} of 110 Å. The number of bonds remains constant at 110 for the remainder of the reaction coordinate.

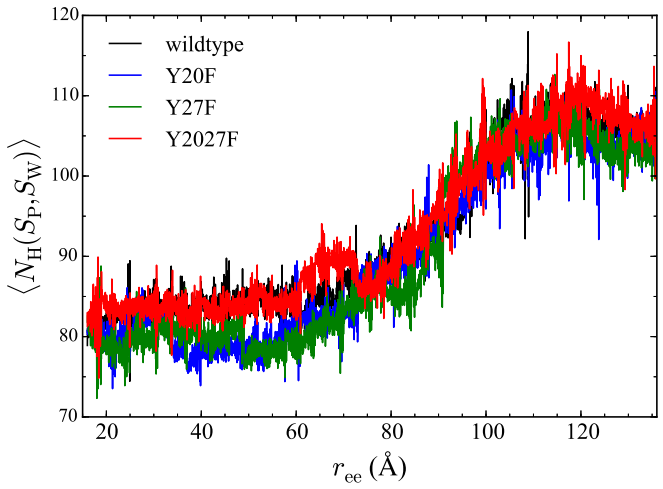

Figure 59: The weighted average hydrogen bonds formed between the peptide and explicit water solvent for the WT NPY (black), Y20F (blue), Y27F (green), and Y2027F (red). The average is of 100 trajectories per stage at pulling velocity of 10 Å/ns.

One interesting question to investigate is whether or not there is a significant number of hydrogen bonds forming between the helix and unstructured tail region of the peptide. This particular question is answered in Fig. 60. In general, each peptide forms no bonds in this region. The WT peptide transiently forms 3 such bonds at a r_{ee} of 75 Å. Similarly, the Y20F mutant forms 3 peptide-solvent bonds at $\approx$ a r_{ee} of 95 Å. Overall, there are very few bonds formed between the helix and the tail during the mechanical unfolding of the peptides. This suggests that all of the intrapeptide hydrogen bonds observed in Fig. 58 are formed only within the helix.

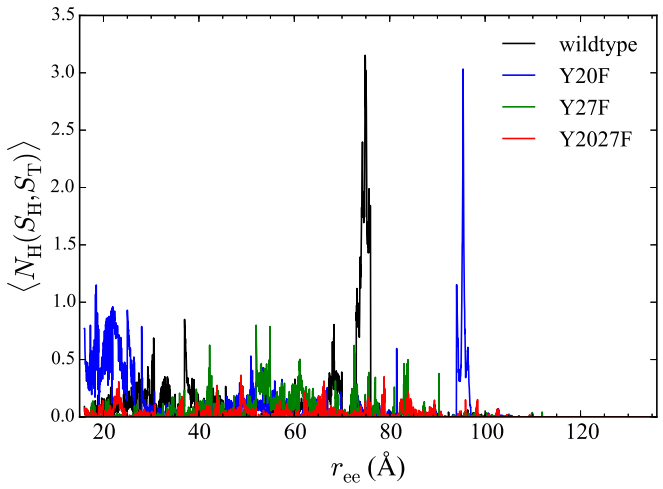

Figure 60: The weighted average hydrogen bonds formed between the helix region of the peptide and the unstructured "tail" for the WT NPY (black), Y20F (blue), Y27F (green), and Y2027F (red). The average is of 100 trajectories per stage at pulling velocity of 10 Å/ns.

5.3.1.4 Thermal unfolding

Complementary to the mechanical unfolding simulations of the individual peptides, far UV CD spectra yields information on the type of secondary structure motifs occurring within the peptide. The far UV spectra shown in Fig. 61 indicates that the peptides are all stable α-helices. This is consistent with previous CD spectra [232, 233, 239]. The mutations at the 20 and 27th positions do not have any effect of the overall secondary structure obtained by the peptides. This is also consistent with the simulation results of the hydrogen bonding profiles described in Sec. 5.3.1.3. The thermal unfolding of the peptides can also be extracted from the CD measurements. Fig. 62 shows the thermal unfolding spectra of the WT and three mutants. The melting temperature, T_m, is the temperature at which 50% of the peptide become unfolded. WT NPY (black curve) has a T_m of $\approx$ 38°C The T_m of the double mutant, Y2027F, is shifted by 30°C in comparison to the WT unfolding. The T_m of the single mutants, Y20F and Y27F, are the same at $\approx$ 60 °C. From the results, a trend in the stability of the peptides begins to emerge: the double mutant is the most thermally

stable, the WT is the least stable and the single mutants are the more stable than the WT but less stable than the double mutant. Though the mutations do not effect the secondary structure of the peptide, the increased packing of the hydrophobic region between the helix and the unstructured tail of the peptide serves to increase its thermal stability.

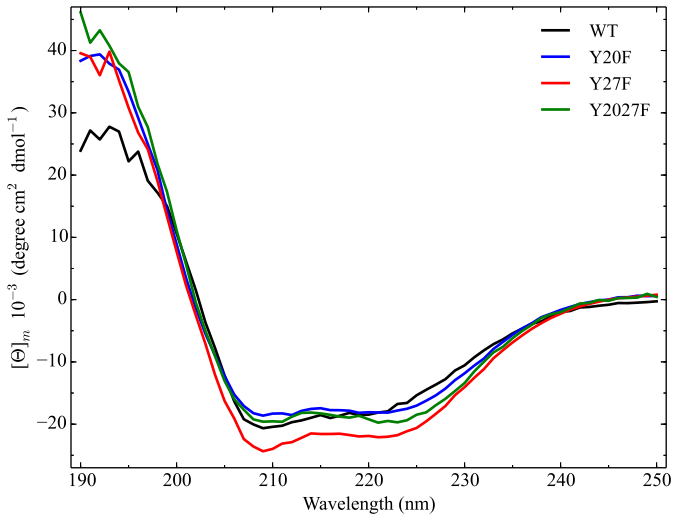

Figure 61: Far UV CD spectra of WT NPY (black), Y20F (blue), Y27F (red), and Y2027F (green). This data was contributed by Dr. Stephen Quirk.

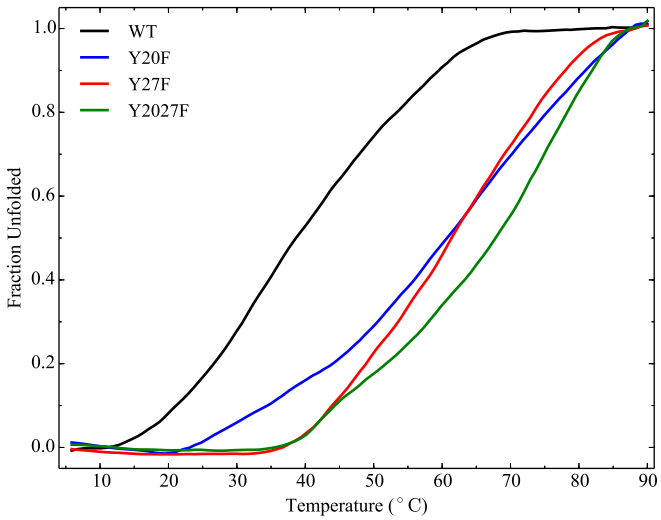

Figure 62: Thermal denaturation of WT NPY and mutants as measured by CD spectroscopy at 222 nm. The color scheme is the same as Fig. 61. Mean molar ellipticity values have been converted to the fraction unfolded peptide. This data was contributed by Dr. Stephen Quirk.

5.3.1.5 Receptor binding

The next aspect of our investigation focuses on the effects the mutations have on the function and behavior of the peptides. In order for NPY to reach a receptor, it must first transverse a lipid membrane. The binding WT NPY and the three mutants to a lipid membrane observed using Surface Plasmon Resonance (SPR) are shown in Fig. 63. SPR is a biophysical technique used to capture protein-protein and protein-membrane interactions [240]. In an SPR experiment, a ligand is attached to a gold surface while an analyte is pumped across the surface. A polarized laser is directed at the gold surface. The gold surface will generate surface plasmons when hit with the laser at a specific angle. The critical angle depends on the refractive index and changes as analyte binds to the ligand. SPR is reported in"Resonance Units" which correspond to a critical angle shift when a ligand binds with an analyte. In this experiment, the ligand is a POPC/POPG lipid vesicle and the analyte is NPY. The WT NPY (black) has the fastest and strongest binding to the lipid membrane.

The single mutant, Y20F, has the next fastest binding rate followed by Y27F. Y2027F has the most inhibited binding to the lipid membrane. These results suggest the the mutations substantially effect the ability of the peptide to bind to a lipid membrane. This is significant as part of the reason NPY has been studied heavily in the literature as a potential drug target is because of its ability to easily transverse the blood-brain barrier.

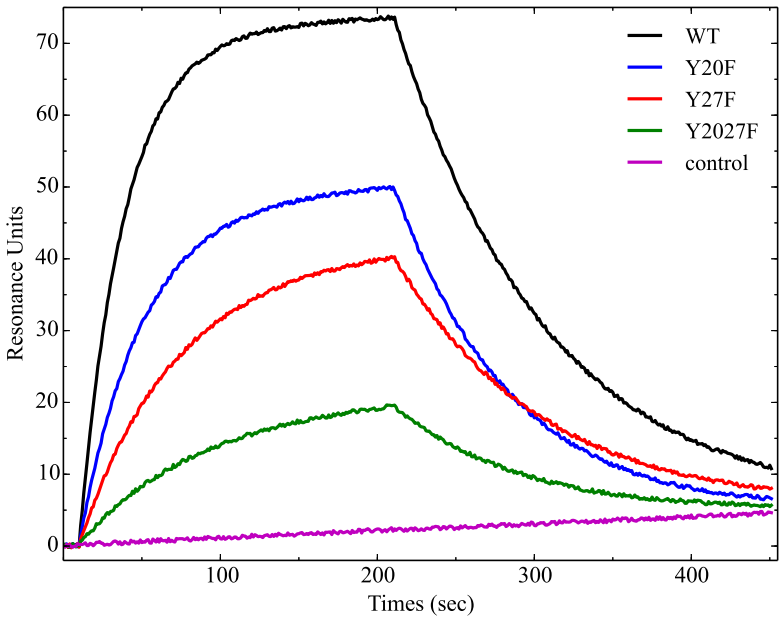

Figure 63: Surface plasmon resonsance of WT NPY and mutants. The color scheme is the same as Fig. 61. This data was contributed by Dr. Stephen Quirk.

In order to further assess the effects that the mutations have on the function of NPY, a competitive binding assay was performed. A fluorescent tag, Pacific Blue, was attached to the NPY peptides. Rat brain homogenate was used as the source of the receptors as it contains a mixture of Y receptor subtypes to which NPY is known to bind [241]. The half-maximal inhibitory concentration, IC_{50}, was measured. This is yields information on how effectively a substance inhibits a specific or biochemical function. The WT peptide has an IC_{50} of 1.3±0.6 nM, Y20F has the next smallest IC_{50} at 1.4±0.4 nM, followed by Y27F at 2.2±0.5 and Y2027F with 4.1±0.6. This trend matches the trend in stability observed

in the SPR experiment in Fig. 63. Overall, the mutants had no effect on the ability of the peptides to bind to the receptor.

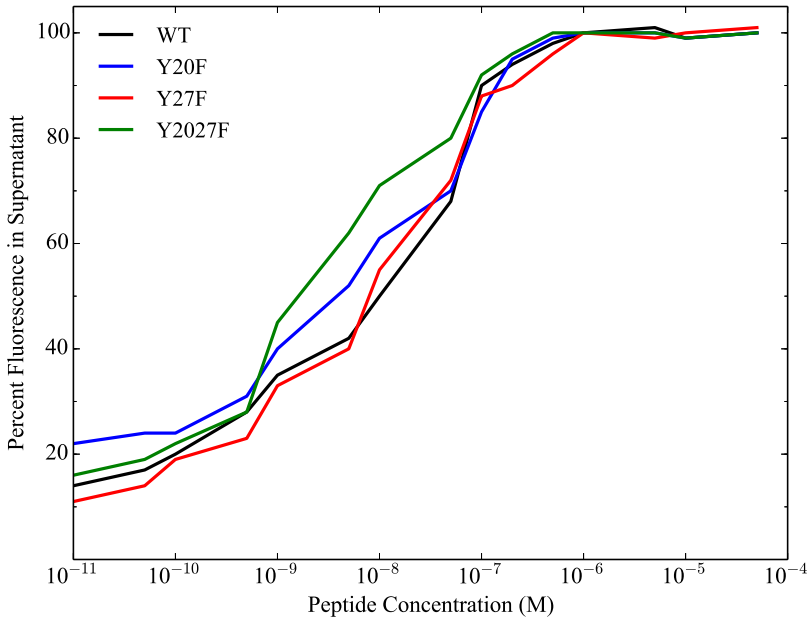

Figure 64: Release of receptor-bound Pacific Blue-labeled WT NPY and mutants upon the addition of unlabeled competitor peptide. The color scheme is the same as Fig. 61 The percent fluorescence in the supernatant is plotted as a function of competitor peptide concentration. This data was contributed by Dr. Stephen Quirk.

5.3.1.6 Analytical ultracentrifugation

As it was previously discovered that NPY can form dimers in solution [226, 242, 243], it is useful to investigate whether or not the mutant NPY peptides in this study can form stable higher order structures, as well. In order to examine this aspect of the structure, analytical ultracentrifugation (AUC) was used. AUC experiments rely on the mass of the molecules of interest and the law of gravitation [244]. A molecular solution is spun at a very high velocity. The sedimentation rate of the different species in the solution can be differentiated to determine if any concentration dependent association occurred. The sedimentation coefficient, s, is a ratio of the velocity to the centrifugal field applied to the

solution. The results of an AUC experiment using WT NPY (a), Y20F (b), Y27F (c), and Y2027F (d) are shown in Fig. 65. These experiments are performed in water at 20°C to mimic the biological environment and at four different concentrations per peptide. The concentrations increase from top to bottom for each peptide.

The observed size distribution of WT varies as a function of concentration. At low concentration (0.0281 mg/mL), a single predominant species with a molecular weight equal to 4.92 kDa is observed. Since the molecular weight of monomeric WT peptide based on amino acid sequence is 4.27 kDa, this peak most likely represents monomeric WT peptide. Therefore, WT peptide appears to exist solely as monomer at 0.0281 mg/mL. Conversely, the analysis at intermediate concentrations of 0.145 mg/mL and 0.305 mg/mL suggest two sedimenting species. Namely, the smallest species has a molecular weight of 5.50 kDa and 4.65 kDa at 0.145 mg/mL and 0.305 mg/mL, respectively. These species correspond with the single species observed at 0.0281 mg/mL and similarly are most likely monomeric WT peptide. However, the larger species at 0.145 mg/mL has a molecular weight equal to 11.8 kDa. Likewise, the larger species at 0.305 mg/mL has a molecular weight equal to 11.1 kDa These species, as well as the intermediate species observed at 0.903 mg/mL with a molecular weight equal to 16.9 kDa represent a common oligomer consisting of a di-, tri-, or tetramer of WT peptides. These results, in conjunction with the absence of any oligomer at the lowest concentration of WT peptide, suggest a concentration dependent reversible self-association (RSA) of WT peptides. Moreover, analysis at 0.903 mg/mL produces a third species with a molecular weight of 29.3 kDa in addition to monomer and the intermediate oligomer species. Based on molecular weight analysis, it is most likely a hexa- or heptamer of WT peptides. Only the highest concentration of WT peptide displays the third species, which suggests that the formation of this species is concentration dependent as well and that the formation of the intermediate oligomer may not represent the extent of NPY RSA.

The mutations to the WT peptide appear to shift the size distribution of species to the right. That is, the mutations promote RSA and the formation of the intermediate and third oligomeric species to various degrees. For the Y20F peptide, three distinct species sediment at a concentration of 0.0286 mg/mL. The smallest species has a molecular weight of 3.88

kDa, which is indicates a monomer, the intermediate species with a molecular weight of 13.8 kDa is most likely the intermediate oligomer as observed for the WT peptide at several concentrations. The largest species has a molecular weight of 28.1 kDa and is likely similar to the third species observed for the WT peptide at 0.903 mg/mL. However, the WT peptide consists of only one species identified as monomer at a the low concentration. Furthermore, the third species is not seen for the WT peptide until a the high concentration and yet is not the dominate species present. In contrast, for the Y20F and Y27F peptides the third species is the dominate species at more than a three-fold less concentration (0.266 mg/mL). At those concentrations, the monomeric species is no longer the most prevalent species. These results suggest that the Y20F and Y27F mutations alter the stability of the NPY peptide in a manner that promotes RSA. This could be an indication of the importance of Tyr20 and Tyr27 in the stability and formation of higher-order structure of NPY peptide.

Similarly, a slight peak that is most likely monomer is only seen at the lowest concentration of 0.0370 mg/mL for Y2027F. At higher concentrations, only a single oligomeric peptide is observed. It likely corresponds to the third species observed in Y20F and Y27F as it has an average molecular weight of 25.5 kDa. This molecular weight suggests it is a hexa- or hepta-mer. The absence of any intermediate species (di-,tri-, or tetramer) and the dominance of the heavier species suggests that the mutations may be additive with respect to the association of higher-order structures in NPY.

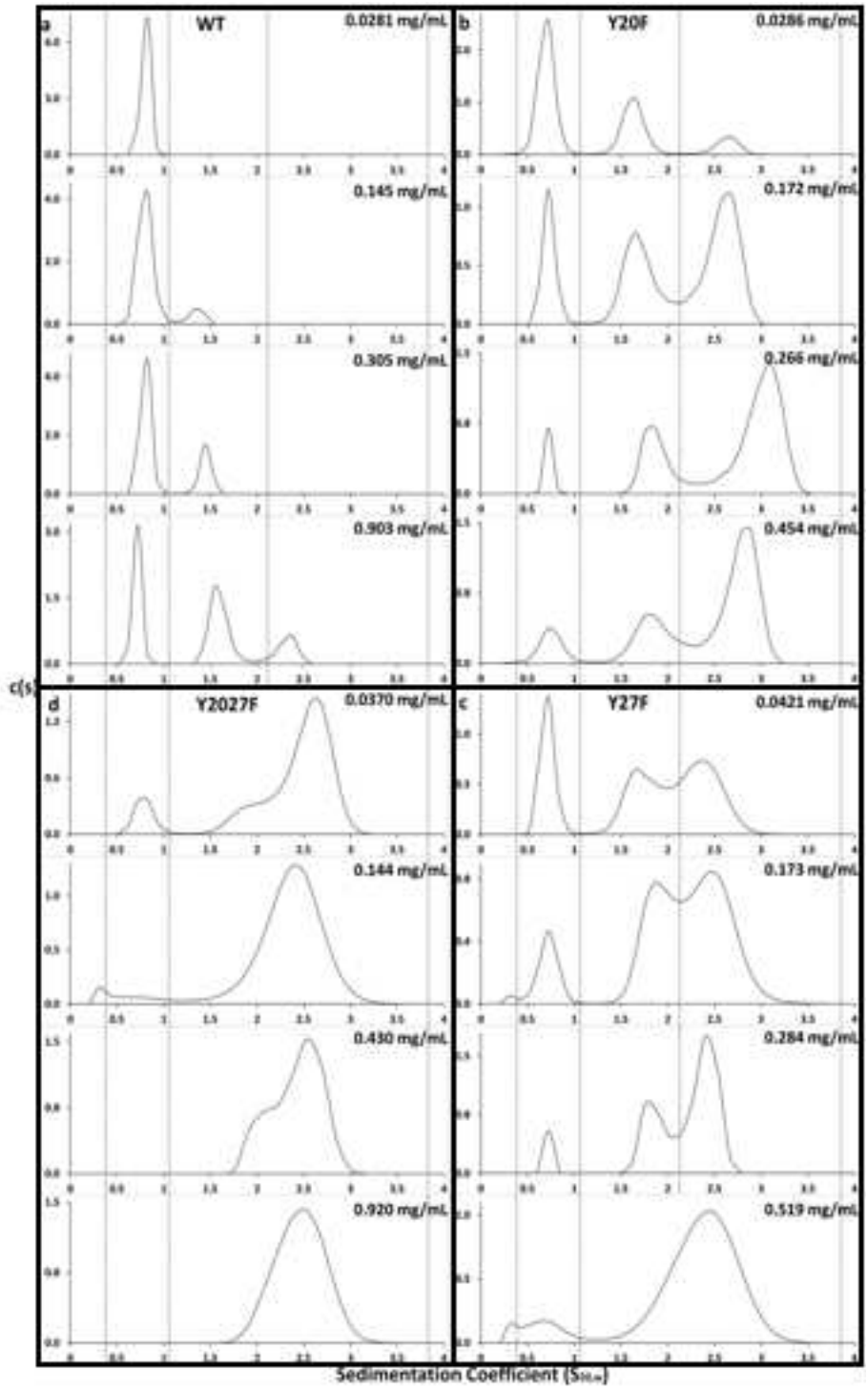

Figure 65: Size distributions of a) WT, b) Y20F, c) Y27F, and d) Y2027F NPY peptides at four concentrations per peptide. The sedimentation coefficients are expressed in water at 20°C. Concentrations are shown in the upper right hand corner of each respective distribution. Dotted lines represent boundaries for similar sedimenting species across concentrations and types of NPY peptides. This data was contributed by Dr. David Bain.

5.3.2 Enterocin 7B

5.3.2.1 Energetics and convergence

The preliminary results of the mechanical unfolding energetics of Enterocin 7B are presented in Fig. 66. Results using seven of the ten mutations suggested in Sec. 5.2.2, namely I4R, A5L, F15R, I23R, F26R, F36L, and L40R have substantial effects on the stability of WT Enterocin 7B. The first point of comparison of the PMFs is the r_{ee} at which the peptides reach a local minimum. The WT (red curve), A5L (blue curve), and F26L (black curve) reach a well-defined local minimum at $\approx$ 21 Å. The I4R mutant (cyan curve) reaches a more compact structure with a local minimum at $\approx$ 18 Å. F26R (magenta curve) has a less compact structure and reaches a well-defined minimum at 24 Å. Mutants F15R (green curve), L40R (gray curve), and I23R (yellow curve) have very broad minimum regions. I23R does exhibit a slight minimum that is shifted to $\approx$ 28 Å, which indicates that the structure has a more relaxed structure than the WT.

Another point of comparison is the overall behavior of the PMFs. The PMFs of the L40R and F26R mutants are most similar to the WT. This indicates that this mutation does not effect the overall unfolding pathway of Enterocin 7B. I23R has a very destabilized PMF in comparison to the WT and any other mutant. The overall energy of unfolding 29 kcal/mol. A5L and F26L have regions of the unfolding coordinate where their stability exceeds that of the WT. The PMF of A5L becomes higher than WT between 30 to 50 Å, while the PMF of F26L surpasses the WT PMF from 22 to 65 Å. However, both A5L and F26L have regions of the PMF where they are less stabilized than the WT. The F15R and I4R mutants both result in PMFs that substantially stabilize the peptide, which was the opposite of what was hypothesized. Their total unfolding energies are $\approx$ 50 kcal/mol while that of the WT is $\approx$ 42 kcal/mol.

The ASMD simulations of the unfolding of L7R, V8R, and M24L are currently underway. It is expected that the results of those simulations will align with and reinforce the findings from the other seven mutations.

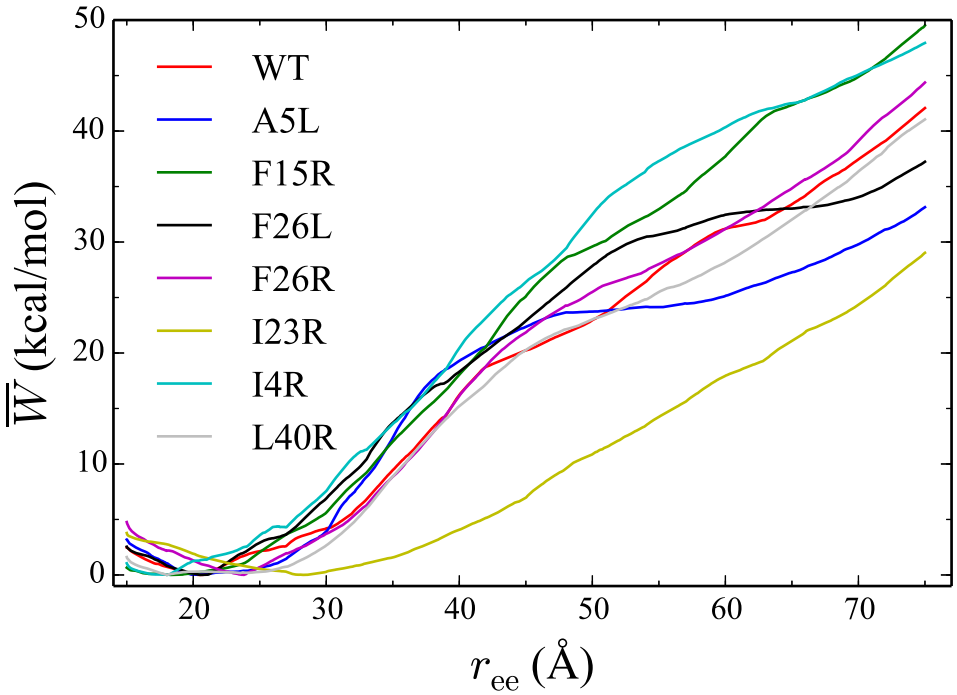

Figure 66: PMFs obtained for all for peptides using a sampling size of 100 trajectories per stage at a pulling velocity of 1 Å/ns using explicit water solvent.

5.4 Conclusions

5.4.1 NPY mutations

In this study, we have combined computational and experimental techniques to investigate the effect three individual mutations have on the stability of the hydrophobic core of NPY. We use a combination of techniques in order to span the scale of investigation from single-molecule dynamics to formation of quaternary structure. In the WT NPY structure, a hydrophobic pocket exists between the helical region and the unstructured tail. Hydrophobic residues, namely proline, are responsible for this core region. Tyrosine residues occupy the 20th and 27th positions within the sequence (within the helical portion of the peptide) and as tyrosine is a polar residue, it does not contribute to the packing of the hydrophobic core. However, further "packing" of this hydrophobic core has led to interesting results that effect the behavior of the peptide on several different levels.

Computationally, we used ASMD to investigate peptides on the single-molecule scale. In these simulations, the peptides are directed along a stretching pathway from which we can extract the energetics of unfolding, energy of specific interactions and hydrogen bonding patterns that play a role in the stability of the peptides. Our results suggest that there may be two possible unfolding pathways that are slight different from one another. The single mutants result in PMFs that fall within the range of thermal fluctuation of one another and are different than the PMFs obtained for WT and Y2027F. Additionally, from our analysis we gained insight into the stability of the hydrophobic core and helical region. Through the calculation of the interaction energy between the hydrophobic residues within the peptide and the water solvent, we can determine that mutations do effect the r_{ee} at which the hydrophobic core opens. As expected, the WT hydrophobic core opens at a r_{ee} of 60 Å, which indicates that it has the least stabilized core. This is followed by Y20F and Y27F, which both open at 75 Å, and Y2027F opening at 80 Å. Based on our analysis of the hydrogen bond patterns that emerge during the unfolding, only the helical region of the peptide contains intrapeptide hydrogen bonds which further indicates that only hydrophobic packing is contributing to the stability of the region between the helix and tail.

Several different types of experimental techniques have been used in this study to uncover more details about the stability and structure of the peptides. First, through CD measurements it was determined that the mutations do not adversely affect the secondary structure of the peptides. However, in the thermal unfolding experiments, the double mutant has a shifted T_m of $\approx 30^{\circ}$C. From this, it is clear that while the secondary structure is not affected, the packing of the hydrophobic core increases the thermal stability substantially. As the peptide must transverse a lipid membrane in order to reach its intended receptor, it is imperative to study the effect the mutations might have on the interaction the peptide with a membrane. As evidenced by SPR experiments, the mutations do effect the interaction of the peptide with the membrane. The WT binds in the fastest while Y2027F binds the slowest. This indicates that the increase in core stability may be hindering the function of the peptide binding to the membrane. Through our receptor binding study it was found that despite the decreased interaction the mutants have with the membrane, the

rate of binding to the target receptors is not affected at all. In order to further investigate the reasons behind the hindrance of the double mutant to bind to the membrane, AUC experiments were performed to identify if any oligomerization was occurring. These findings overwhelming suggest that Y2027F stability arises due to the formation of higher order structures, specifically as hexa- or heptamers. As the peptide is known to interact with the membrane as a monomer, it is conceivable that the dissociation of the hexa- or heptamer impedes the ability of Y2027F to interact with the lipid membrane.

5.4.2 Enterocin 7B mutations

The goal of this study is to investigate the stability of the hydrophobic core of Enterocin 7B. From preliminary ASMD simulations, it is possible to confirm the overall effect each mutation has on the unfolding pathway of Enterocin 7B. Of the mutants that were proposed as destabilizing only I23R can be confirmed to destabilize the structure. Surprisingly, mutants I4R and F15R, which were predicted to be destabilizing are actually stabilizing. A5L and F26L, which were hypothesized to act as peptide stabilizers, do indeed add a measure of stability to the structure in the mid-region of the unfolding pathways. L40R and F26R which were proposed as a destabilizing mutations, do not effect the unfolding pathway of Enterocin 7B.

However, as the theoretical study is still currently underway, there is much left to learn about the factors that effect the stability of Enterocin 7B. Analysis of the hydrogen bond patterns and residue pair interaction energies are expected to give greater insight into the reasons behind some of the surprising preliminary results. Additionally, experimental studies of how the mutations effect the antimicrobial properties of Enterocin 7B will also yield important information about the overall effects the mutations have on the peptide.

CHAPTER VI

CONCLUSIONS AND OUTLOOK

6.1 Efficacy of ASMD as an enhanced-sampling method

ASMD has been used to efficiently calculate the PMFs for the mechanical unfolding of several different peptides. The results of the converged PMFs are equivalent to those obtained using SMD though are achieved using an order of magnitude less calculation time and orders of magnitude faster pulling velocity. ASMD is able to overcome known sampling limitations that accompany traditional SMD by implementing SMD in a segmented approach using short non-equilibrium stretching stages that are alternated with contraction stages. As such several variations of ASMD have been developed in order to test the accuracy and robustness of the methodology.

In naïve ASMD, the contraction stage is the weighted Jarzynski Average is used as a metric by which to select a trajectory to act as a seed for the next stage of the calculation. This approach has been benchmarked using helical and β-hairpin peptides as presented in Chapters 3 and 4. However, one potential pitfall of naïve ASMD is that the projections are skewed towards a single dominant pathway. For small peptides, this may not be such a problem as there may be only one dominant unfolding pathway by which the peptide can unfold. However, in larger peptide systems this possibility is reduced as the the number of possible conformations is increased due to the increase in the number of degrees of freedom. Three additional methods based on ASMD have been developed, namely, FR-ASMD, PR-ASMD, and MB-ASMD, to avoid such biasing. The intermediate stages for contraction used in FR-ASMD and PR-ASMD perform additional equilibrations and minimizations which help to avoid biasing as there is no selection criterion used to evaluate the trajectories. This effectively allows for multiple pathways to be observed. However, the addition of the equilibration stages adds back significant computational time which defeats the efficiency advantage of using ASMD. FR-ASMD has been seen to provide converged

PMFs of ALA_{10} in vacuum, implicit solvent, and explicit solvent and using a wide range pulling velocities. Future work will use and benchmark PR-ASMD for the simulation of larger peptide systems. MB-ASMD using a probabilistic approach for the contraction stage to select an ensemble of trajectories with which to continue the simulations. Thus far, this variation of ASMD has been seen to provide very accurate PMFs using ALA_{10} in vacuum conditions as the benchmark system. Though there remains benchmarking that can be performed with FR-ASMD, PR-ASMD, and MB-ASMD, the methods are robust and promise the ability to converge PMFs of biomolecular processes on a much faster timescale. We anticipate that these methods will allow for simulation of larger and more complex systems while maintaining computational tractability

6.2 Creating a library of the mechanical unfolding energetics of small, secondary structure motifs

A central achievement of this thesis work is the advancement of the understanding of the dynamics and energetics of peptides in complex biological environments. In this work, we have covered a broad range of secondary structure motifs including different types of helices, β-hairpins, and mixed-motif peptides such as polyproline helices and helical bundles. By creating a small library of the mechanical unfolding pathways of a wide variety of peptides, we have the ability to unravel the links between their structure, dynamics, and function. The addition of these unfolding pathways to the standard repertoire of data associated with specific peptide sequences (i.e. secondary structure, CD spectra, melting temperature, kinetics of binding, etc.) has the potential to transform the characterization of their dynamic behavior.

The level of detail gained through our MD simulations allows us to capture information about the unfolding pathway. For example, we calculate the hydrogen bonds that form and break over the course of the unfolding as well as the interaction energies of specific residue pairs that contribute to the stabilization or destabilization of the peptide. This high-resolution information gives use much greater insight into the causes of a peptide's behavior. Several of our studies also include original mutations of the WT peptides which allow us to probe the resiliency of the secondary structure stability. When considered

together, this work offers an additional viewpoint from which to approach solving problems in chemical biology.

6.3 Using experimental data as a correlative tool

As the challenges facing chemical biology become increasingly complex, it requires a multi-disciplinary approach to solve such problems. MD is designed to capture the atomistic scale of molecules undergoing dynamic transition whereas most experimental methods capture macro-scale observations associated with molecular interactions. As such, experiment and computation offer the perfect complement to one another. Biomolecular systems are, in many ways, the ideal systems to use for joint experimental and computational studies since numerous methods have been developed to characterize their behavior. The development and availability of accurate FFs that describe proteins is a very active area of research within MD and experimental methods to address problems in chemical biology are many and varied.

This thesis has taken advantage of combining both areas of research. Two of our studies found in Chapters 4 and 5 rely heavily of the use of experimental methods to clarify and add depth to the studies. In the first study, we successfully correlate the mechanical unfolding of several mutations of trpzip1 with experimental thermal unfolding. From this, we can extract information about which mutations result in stabilizing trpzip1 and which mutations destabilize the peptide. We also combined theoretical and experimental methods to study the effects of mutations on NPY. We were able to span the scale of investigation from the dynamics of a single molecule to the formation of higher-order structures. From this, we concluded that a double mutation that serves to more tightly pack the hydrophobic core of NPY also causes changes to the ability of the peptide to traverse the lipid membrane.

MD can be used as a predictive tool to direct design of biomolecules in cases where the experiments are too costly or the system does not easily lend itself to experimentation. The synergy of combining computational and experimental methods to approach problems in chemical biology is just beginning to take hold. The use of ASMD in combination with experimental studies has led to the discovery of novel peptide sequences whose dynamics,

structure, and function have been well-characterized.

6.4 Toward using ASMD to simulate more complex systems

In principle, the ASMD methodology could be applied to any system which undergoes a transition between non-equilibrium states. Therefore, it could potentially be a useful tool to study biophysical problems beyond peptide unfolding. In the same way that SMD was used to study a variety of problems such as ion channel transport and dissociation of a salt as presented in Chapter 2, ASMD and the variations of ASMD could be used to tackle similar problems. One avenue of investigation ASMD could be used to explore is the simulation of protein-surface interactions. For example, this could include protein-protein, protein-membrane surface, or protein-nanoparticle interactions. It could also be used to simulate the dynamic motions of a protein translocating through a lipid membrane.

The simulation of ligand binding events would also be one potentially very useful application of ASMD. However, the inherent linearity associated with steered unfolding simulations would need to be overcome in order to simulate such systems as their reaction coordinates would most likely be non-linear. For example, incorporation of non-linear reaction coordinates for generating 2D PMFs would be required if the goal of the simulations were to provide clarity on the native behavior of such processes.

REFERENCES

[1] Levinthal, C., Are there pathways for protein folding?, *J. Chim. Phys.*, vol. 65, pp. 44–45, 1968.

[2] Bryngelson, J. D. and Wolynes, P. G., Intermediates and Barrier Crossing in a Random Energy Model (with Applications to Protein Folding), *J. Phys. Chem.*, vol. 93, pp. 6902–6915, 1989.

[3] Wolynes, P. G., Symmetry and the Energy Landscapes of Biomolecules, *Proc. Natl. Acad. Sci. U.S.A.*, vol. 93, pp. 14249–14255, 1996.

[4] Mayor, U., Johnson, C. M., Daggett, V., and Fersht, A. R., Protein folding and unfolding in microseconds to nanoseconds by experiment and simulation, *Proc. Natl. Acad. Sci. U.S.A.*, vol. 97, pp. 13518–13522, 2000.

[5] Daggett, V. and Fersht, A. R., Is there a unifying mechanism for protein folding?, *Trends Biochem. Sci.*, vol. 28, pp. 18–25, 2003.

[6] Selkoe, D. J., Folding proteins in fatal ways, *Nature (London)*, vol. 426, no. 6968, pp. 900–904, 2003.

[7] Binnig, G., Quate, C., and Gerber, C., Atomic force microscope, *Phys. Rev. Lett.*, vol. 56, no. 9, pp. 930–933, 1986.

[8] Chen, J., Lu, Z., Sakon, J., and Stites, W. E., INCREASING THE THERMOSTABILITY OF STAPHYLOCOCCAL NUCLEASE: IMPLICATIONS FOR THE ORIGIN OF PROTEIN THERMOSTABILITY, *J. Mol. Biol.*, vol. 303, pp. 125–130, 2000.

[9] Best, R. B., Li, B., Steward, A., Daggett, V., and Clarke, J., Can non-mechanical proteins withstand force? Stretching barnase by atomic force microscopy and molecular dynamics simulation, *Biophys. J.*, vol. 81, pp. 2344–2356, 2001.

[10] Schuler, B., Lipman, E. A., and Eaton, W., Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy, *Nature (London)*, vol. 419, no. 6908, pp. 743–747, 2002.

[11] Butt, H. J., Cappella, B., and Kappl, M., Force measurements with the atomic force microscope: Technique, interpretation and applications, *Surf. Sci. Rep.*, vol. 59, no. 1-6, pp. 1–152, 2005.

[12] Pusey, M. L., Liu, Z. J., Tempel, W., Praissman, J., Lin, D., Wang, B. C., Gavira, J. A., and Ng, J. D., Life in the fast lane for protein crystallization and X-ray crystallography, *Prog. Biophys. Mol. Biol.*, vol. 88, pp. 359–386, 2005.

[13] Swope, W., Pitera, J., and Suits, F., Describing Protein Folding Kinetics by Molecular Dynamics Simulations. 1. Theory, *J. Phys. Chem. B*, vol. 108, pp. 6571–6581, 2004.

[14] Swope, W., Pitera, J., Suits, F., Eleftherious, M., Fitch, B., Germain, R., Rayshubski, A., Ward, T., Zhestkov, Y., and Zhou, R., Describing Protein Folding Kinetics by Molecular Dynamics Simulations. 2. Example Applications to Alanine Dipeptide and Beta-Hairpin Peptide, *J. Phys. Chem. B*, vol. 108, pp. 6582–6594, 2004.

[15] Halperin, I., Ma, B., Wolfson, H., and Nussinov, R., Principles of docking: an overview of search algorithms and a guide to scoring functions, *Proteins*, vol. 47, pp. 409–443, 2002.

[16] Okur, A., Roe, D. R., Cui, G., Hornak, V., and Simmerling, C., Improving Convergence of Replica-Exchange Simulations through Coupling to a High-Temperature Structure Reservoir, *J. Chem. Theory Comput.*, vol. 3, pp. 557–568, 2007.

[17] Anfinsen, C. B., Principles that govern folding of protein chains, *Science*, vol. 181, no. 4096, pp. 223–230, 1972.

[18] Ashkin, A. and Dziedzic, J., Optical trapping and manipulation of viruses and bacteria, *Science*, vol. 235, no. 4795, pp. 1517–1520, 1987.

[19] Evans, E., Ritchie, K., and Merkel, R., Sensitive Force Technique to probe molecular adhesion and structural linkages at biological interfaces, *Biophys. J.*, vol. 68, no. 6, pp. 2580–2587, 1995.

[20] Lu, H., Isralewitz, B., Krammer, A., Vogel, V., and Schulten, K., Unfolding of Titin Immunoglobulin Domains by Steered Molecular Dynamics Simulation, *Biophys. J.*, vol. 75, pp. 662–671, 1998.

[21] Lu, H. and Schulten, K., Steered molecular dynamics simulations of force-induced protein domain unfolding, *Proteins*, vol. 35, pp. 453–463, 1999.

[22] Isralewitz, B., Baudry, J., Gullingsrud, J., Kosztin, D., and Schulten, K., Steered Molecular Dynamics Investigations of Protein Function, *Journal of Molecular Graphics and Modelling*, vol. 19, pp. 13–25, 2001.

[23] Park, S., Khalili-Araghi, F., Tajkhorshid, E., and Schulten, K., Free Energy Calculation from Steered Molecular Dynamics Simulations Using Jarzynski's Equality, *J. Chem. Phys.*, vol. 119, pp. 3559–3566, 2003.

[24] Park, S. and Schulten, K., Calculating Potentials of Mean Force from Steered Molecular Dynamics Simulations, *J. Chem. Phys.*, vol. 120, pp. 5946–5961, 2004.

[25] Kim, I., Xu, W., and Reed, J. C., Cell death and endoplasmic reticulum stress: disease relevance and therapeutic opportunities, *Nat. Rev. Drug Discovery*, vol. 7, no. 12, pp. 1013–1030, 2008.

[26] Ross, C. A. and Poirier, M. A., Protein Aggregation and Neurodegenerative Disease, *Nautre Medicine*, vol. 10, pp. S10–S17, 2004.

[27] Bureau, H. R., Merz Jr., D., Hershkovits, E., Quirk, S., and Hernandez, R., Constrained Unfolding of a Helical Peptide: Implicit Versus Explicit Solvents, *PLoS ONE*, vol. 10, p. e0127034, 2015.

[28] Allen, C., Bureau, H. R., Quirk, S., and Hernandez, R. "Direct comparison of the energetics of decaalanine using CHARMM force field variants," in preparation.

[29] Bureau, H. R., Jr., T. D. M., Allen, C., Quirk, S., and Hernandez, R. "Comparison of Amber16 and CHARMM36 force fields using two different secondary structure motifs," in preparation.

[30] Bureau, H. R., Bucher, R., Quirk, S., and Hernandez, R. "Energetics and structure of alanine-rich α-helices," in preparation.

[31] Bureau, H. R., Hershkovits, E., Quirk, S., and Hernandez, R., Determining the energetics of small beta sheet peptides using Adaptive Steered Molecular Dynamics, *J. Chem. Theory Comput.*, vol. 12, pp. 2028–2037, 2016.

[32] Bureau, H. R., Quirk, S., and Hernandez, R. "Correlating the experimental and computational effects of mutations in trpzip1," in preparation.

[33] Quirk, S., Lusk, R., Bureau, H. R., Allen, C., Hernandez, R., and Bain, D. "Unusual thermal stability in Neuropeptide-Y mutants affect quaternary structure and biological activity," in preparation.

[34] Quirk, S., Allen, C., Bureau, H., and Hernandez, R. "Structure, energetics and stability of Enterocin 7B mutants," in preparation.

[35] Zwanzig, R. W., High temperature equation of state by a perturbation method. i. nonpolar gases, *J. Chem. Phys.*, vol. 22, pp. 1420–1426, 1954.

[36] Torrie, G. M. and Valleau, J. P., Nonphysical Sampling Distributions in Monte Carlo Free-Energy Estimation: Umbrella Sampling, *J. Chem. Phys.*, vol. 23, no. 2, pp. 187–199, 1977.

[37] Kumar, S., Bouzida, D., Swendsen, R. H., Kollman, P. A., and Rosenburg, J. M., The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method, *J. Comput. Chem.*, vol. 13, pp. 1011–1021, 1992.

[38] Ferrenberg, A. M. and Swendsen, R. H., New Monte Carlo technique for studying phase transitions, *Phys. Rev. Lett.*, vol. 61, no. 23, pp. 2635–2638, 1988.

[39] Ferrenberg, A. M. and Swendsen, R. H., Optimized Monte Carlo data analysis, *Phys. Rev. Lett.*, vol. 63, no. 12, pp. 1195–1198, 1989.

[40] Tajkorshid, E., Nollert, P., Jensen, M. O., Miercke, L. J. W., O'Connell, J., Stroud, R. M., and Schulten, K., Control of the selectivity of the aquaporin water channel family by global orientational tuning, *Science*, vol. 296, no. 5567, pp. 525–530, 2002.

[41] Gao, M., Craig, D., Vogel, V., and Schulten, K., Identifying unfolding intermediates of FN-III10 by steered molecular dynamics, *J. Mol. Biol.*, vol. 323, pp. 939–950, 2002.

[42] Torras, J., de M. Seabra, G., and Roitberg, A. E., A Multiscale Treatment of Angeli's Salt Decomposition, *J. Chem. Theory Comput.*, vol. 5, pp. 37–46, 2009.

[43] Jarzynski, C., Equilibrium free-energy differences from nonequilibrium measurements: A master-equation approach, *Phys. Rev. E*, vol. 56, pp. 5018–5035, 1997.

[44] Jarzynski, C., Nonequilibrium equality for free energy differences, *Phys. Rev. Lett.*, vol. 78, pp. 2690–2693, 1997.

[45] Jarzynski, C., Equilibrium free-energy differences from nonequilibrium measurements: A master-equation approach, *Phys. Rev. E*, vol. 56, pp. 5018–5035, 1997.

[46] Crooks, G. E., Nonequilibrium measurements of free energy differences for microscopically reversible Markovian systems, *J. Stat. Phys.*, vol. 90, pp. 1481–1487, 1998.

[47] Wu, D. and Kofke, D. A., Phase-space overlap measures. I. Fail-safe bias detection in free energies calculated by molecular simulation, *J. Chem. Phys.*, vol. 123, p. 054103, 2005.

[48] Wu, D. and Kofke, D. A., Phase-space overlap measures. II. Design and implementation of staging methods for free-energy calculations, *J. Chem. Phys.*, vol. 123, p. 084109, 2005.

[49] Moradi, M., Babin, V., Roland, C., and Sagui, C., A classical molecular dynamics investigation of the free energy and structure of short polyproline conformers, *J. Chem. Phys.*, vol. 133, no. 12, p. 125104, 2010.

[50] Wu, D. and Kofke, D. A., Rosenbluth-sampled nonequilibrium work method for calculation of free energies in molecular simulation, *J. Chem. Phys.*, vol. 122, p. 204104, 2005.

[51] Lu, N. and Kofke, D. A., Optimal intermediates in staged free energy calculations, *J. Chem. Phys.*, vol. 111, pp. 4414–4423, 1999.

[52] Echeverria, I. and Amzel, L. M., Helix propensities calculations for amino acids in alanine based peptides using Jarzynski's equality, *Proteins: Struct., Func., Bioinf.*, vol. 78, pp. 1302–1310, 2010.

[53] Echeverria, I. and Amzel, L. M., Estimation of Free-Energy Differences from Computed Work Distributions: An Application of Jarzynski's Equality, *J. Phys. Chem. B*, vol. 116, no. 36, pp. 10986–10995, 2012.

[54] Ozer, G., Valeev, E., Quirk, S., and Hernandez, R., Adaptive Steered Molecular Dynamics of the Long-Distance Unfolding of Neuropeptide Y, *J. Chem. Theory Comput.*, vol. 6, pp. 3026–3038, 2010.

[55] Ozer, G., Quirk, S., and Hernandez, R., Adaptive Steered Molecular Dynamics: Validiation of the Selection Criterion and Benchmarking Energetics in Vacuum, *J. Chem. Phys.*, vol. 136, p. 215104, 2012.

[56] Ozer, G., Quirk, S., and Hernandez, R., Thermodynamics of Decaalanine Stretching in Water Obtained by Adaptive Steered Molecular Dynamics Simulations, *J. Chem. Theory Comput.*, vol. 8, pp. 4837–4844, 2012.

[57] MacFadyen, J. and Andricioaei, I., A skewed-momenta method to efficiently generate conformational-transition trajectories, *J. Chem. Phys.*, vol. 123(7), p. 074107, 2005.

[58] Oberhofer, H., Dellago, C., and Boresch, S., Single molecule pulling with large time steps, *Phys. Rev. E*, vol. 75, p. 061106, 2007.

[59] Do, T. N., Carloni, P., Varani, G., and Bussi, G., RNA/Peptide Binding Driven by Electrostatics—Insight from Bidirectional Pulling Simulations, *J. Chem. Theory Comput.*, vol. 9, pp. 1720–1730, 2013.

[60] Paramore, S., Ayton, G. S., and Voth, G. A., Extending the Fluctuation Theorem to Describe Reaction Coordinates, *J. Chem. Phys.*, vol. 126, p. 051102, 2007.

[61] Ozer, G., Keyes, T., Quirk, S., and Hernandez, R., Multiple Branched Adaptive Steered Molecular Dynamics, *J. Chem. Phys.*, vol. 141, p. 064101, 2014.

[62] Hummer, G. and Szabo, A., Free Energy Reconstruction from Nonequilibrium Single-Molecule Pulling Experiments, *Proc. Natl. Acad. Sci. U.S.A.*, vol. 98, pp. 3658–3661, 2001.

[63] Humphrey, W., Dalke, A., and Schulten, K., VMD - Visual Molecular Dynamics, *J. Molec. Graphics*, vol. 14, pp. 33–38, 1996.

[64] Roe, D. R., Okur, A., Wickstrom, L., Hornak, V., and Simmerling, C., Secondary structure bias in generalized Born solvent models: comparison of conformational ensembles and free energy of solvent polarization from explicit and implicit solvation, *J. Phys. Chem. B*, vol. 111, pp. 1846–1857, 2007.

[65] Reddy, A. S., Wang, L., Lin, Y.-S., Ling, Y., Chopra, M., Zanni, M. T., Skinner, J. L., and de Pablo, J. J., Solution Structures of Rat Amylin Peptide: Simulation, Theory, and Experiment, *Biophys. J.*, vol. 98, pp. 443–451, 2010.

[66] Reddy, A. S., Wang, L., Singh, S., Ling, Y. L., Buchanan, L., Zanni, M. T., Skinner, J. L., and de Pablo, J. J., Stable and Metastable States of Human Amylin in Solution, *Biophys. J.*, vol. 99, pp. 2208–2216, 2010.

[67] Paci, E. and Karplus, M., Unfolding proteins by external forces and temperature: The importance of topology and energetics, *Proc. Natl. Acad. Sci. U.S.A.*, vol. 97, pp. 6521–6526, 2000.

[68] Best, R. B., Buchete, N.-V., and Hummer, G., Are current molecular dynamics force fields too helical?, *Biophys. J.*, vol. 95, pp. L7–L9, 2008.

[69] Lin, Z., Riniker, S., and van Gunsteren, W. F., Free Enthalpy Differences between alpha-, pi-, and 3(10)-Helices of an Atomic Level Fine-Grained Alanine Deca-Peptide Solvated in Supramolecular Coarse-Grained Water, *J. Chem. Theory Comput.*, vol. 9, pp. 1328–1333, 2013.

[70] Hazel, A., Chipot, C., and Gumbart, J. C., Thermodynamics of deca-alanine folding in water, *J. Chem. Theory Comput.*, vol. 10, no. 7, pp. 2836–2844, 2014.

[71] Feig, M. and Brooks III, C. L., Recent advances in the development and application of implicit solvent models in biomolecule simulations, *Curr. Opin. Struct. Biol.*, vol. 14, no. 2, pp. 217–224, 2004.

[72] Qin, Z. and Buehler, M. J., Structure and dynamics of human vimentin intermediate filament dimer and tetramer in explicit and implicit solvent models, *J. Molec. Mod.*, vol. 17, no. 1, pp. 37–48, 2011.

[73] Zagrovic, B., Sorin, E. J., and Pande, V., beta-hairpin folding simulations in atomistic detail using an implicit solvent model, *J. Mol. Biol.*, vol. 313, no. 1, pp. 151–170, 2001.

[74] Paci, E. and Karplus, M., Forced Unfolding of Fibronectin Type 3 Modules: An Analysis by Biased Molecular Dynamics Simulations, *J. Mol. Biol.*, vol. 288, no. 3, pp. 441–459, 1999.

[75] Ho, B. K. and Agard, D. A., An improved strategy for generating forces in steered molecular dynamics: the mechanical unfolding of titin, e2lip3 and ubiquitin, *PloS One*, vol. 5, no. 9, p. e13068, 2010.

[76] Luchko, T., Gusarov, S., Roe, D. R., Simmerling, C., Case, D. A., Tuszynski, J., and Kovalenko, A., Three-dimensional molecular theory of solvation coupled with molecular dynamics in Amber, *J. Chem. Theory Comput.*, vol. 6, no. 3, pp. 607–624, 2010.

[77] Zhou, R., Free Energy Landscape of Protein Folding in Water: Explicit vs. Implicit Solvent, *Proteins: Struct. Func. Gen.*, vol. 53, no. 2, pp. 148–161, 2003.

[78] Juneja, A., Ito, M., and Nilsson, L., Implicit Solvent Models and Stabilizing Effects of Mutations and Ligands on the Unfolding of the Amyloid beta-Peptide Central Helix, *J. Chem. Theory Comput.*, vol. 9, no. 1, pp. 834–846, 2013.

[79] Chen, J. and Brooks III, C. L., Implicit modeling of nonpolar solvation for simulating protein folding and conformational transitions, *Phys. Chem. Chem. Phys.*, vol. 10, no. 4, pp. 471–481, 2008.

[80] Sugita, Y. and Okamoto, Y., Replica-Exchange Molecular Dynamics Method for Protein Folding, *Chem. Phys. Lett.*, vol. 314, pp. 141–151, 1999.

[81] MacKerell Jr., A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T. K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher, W. E., Roux, B., Schlenkrich, M., Smith, J. C., Stote, R., Straub, J., Watanabe, M., Wiorkiewicz-Kuczera, J., Yin, D., and Karplus, M., All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins, *J. Phys. Chem. B*, vol. 102, no. 18, pp. 3586–3616, 1998.

[82] MacKerell, Jr., A. D., Feig, M., and Brooks III, C. L., Extending the Treatment of Backbone Energetics in Protein Force Fields: Limitations of Gas-Phase Quantum Mechanics in Reproducing Protein Conformational Distributions in Molecular Dynamics Simulations, *J. Comput. Chem.*, vol. 25, no. 11, pp. 1400–1415, 2004.

[83] Huang, J. and MacKerell, Jr., A. D., CHARMM36 All-Atom Additive Protein Force Field: Validation Based Comparison to NMR Data, *J. Comput. Chem.*, vol. 34, pp. 2135–2145, 2013.

[84] Case, D. A., Cheatham, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M., Onufriev, A., Simmerling, C., Wang, B., and Woods, R. J., The Amber biomolecular simulation programs, *J. Comput. Chem.*, vol. 26, no. 16, pp. 1668–1688, 2005.

[85] Salomon-Ferrer, R., Case, D. A., and Walker, R. C., An Overview of the Amber Biomolecular Simulation Package, *WIREs Comput. Mol. Sci.*, vol. 3, pp. 198–210, 2013.

[86] Jorgensen, W. L., Maxwell, D. S., and Tirado-Rives, J., Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids, *J. Am. Chem. Soc.*, vol. 118, no. 45, pp. 11225–11236, 1996.

[87] Kaminski, G. A., Friesner, R. A., Tirado-Rives, J., and Jorgensen, W. L., Evaluation and Reparameterization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides, *J. Phys. Chem. B*, vol. 105, no. 28, pp. 6474–6487, 2001.

[88] Garcia, A. E. and Sanbonmatsu, K. Y., Alpha-helical Stabilization by Side Chain Shielding of Backbone Hydrogen Bonds, *Proc. Natl. Acad. Sci. U.S.A.*, vol. 99, pp. 2782–2787, 2002.

[89] Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C., Comparison of Multiple Amber Force Fields and Development of Imprved Protein Backbone Parameters, *Proteins: Struct., Func., Bioinf.*, vol. 65, pp. 712–725, 2006.

[90] Maier, J., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K., and Simmerling, C., ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB, *J. Chem. Theory Comput.*, vol. 11, pp. 3696–3713, 2015.

[91] Best, R. B., Zhu, X., Shim, J., Lopes, P. E. M., Mittal, J., Feig, M., and Mackerell, Jr, A. D., Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone phi, psi and Side-Chain chi1 and chi2 Dihedral Angles, *J. Chem. Theory Comput.*, vol. 8, pp. 3257–3273, 2012.

[92] Nadzirin, N. and Firdaus-Raih, M., Proteins of Unknown Function in the Protein Data Bank (PDB): An Inventory of True Uncharacterized Proteins and Computational Tools for Their Analysis, *Int. J. Mol. Sci.*, vol. 13, pp. 12761–12772, 2012.

[93] Taylor, K. S., Chin, M. Z., Yang, N. C., and Garavito, R. M., A novel, multilayer structure of a helical peptide, *Protein Sci.*, vol. 5, pp. 414–421, 1996.

[94] Ghosh, T., Garde, S., and Garcia, A. E., Role of Backbone Hydration and Salt-Bridge Formation in Stability of Alpha-Helix in Solution, *Biophys. J.*, vol. 85, no. 5, pp. 3187–3193, 2003.

[95] MacKerell, Jr., A. D., Feig, M., and Brooks III, C. L., Improved Treatment of the Protein Backbone in Empirical Force Fields, *J. Am. Chem. Soc.*, vol. 126, no. 3, pp. 698–699, 2004.

[96] Best, R. B., Mittal, J., Feig, M., and MacKerell, Jr., A. D., Inclusion of Many-Body Effects in the Additive CHARMM Protein CMAP Potential Results in Enhanced Cooperativity of alpha-Helix and beta-Hairpin Formation, *Biophys. J.*, vol. 103, no. 5, pp. 1045–1051, 2012.

[97] Stone, A. J., Intermolecular Potentials, *Science*, vol. 321, pp. 787–789, 2008.

[98] Bartlett, G. J., Choudhary, A., Raines, R. T., and Woolfson, D. N., n –¿ pi* Interactions in Proteins, *Nat. Chem. Biol.*, vol. 6, no. 8, pp. 615–620, 2010.

[99] Morozov, A. V., Tsemekhman, K., and Baker, D., Electron Density Redistribution Accounts for Half the Cooperativity of Alpha Helix Formation, *J. Phys. Chem. B*, vol. 110, pp. 4503–4505, 2006.

[100] Nilsson, L. and Karplus, M., Empirical Energy Functions for Energy Minimization and Dynamics of Nucleic-Acids, *J. Comput. Chem.*, vol. 7, pp. 591–616, 1986.

[101] Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klien, M. L., Comparison of Simple Potential Functions for Simulating Liquid Water, *J. Chem. Phys.*, vol. 79, pp. 926–935, 1983.

[102] Mackerell, A. D., Wiorkiewiczkuczera, J., and Karplus, M., An All-Atom Empirical Energy Function for the Simulation of Nucleic-Acids, *J. Am. Chem. Soc.*, vol. 117, pp. 11946–11975, 1995.

[103] Feig, M., Mackerell, A. D., and Brooks, C. L., Force Field Influence on the Observation of pi-helical Protein Structures in Molecular Dynamic Simulations, *J. Phys. Chem. B*, vol. 107, pp. 2831–2836, 2003.

[104] Buck, M., Bouguet-Bonnet, S., Pastor, R., and Jr., A. D. M., Importance of the CMAP correction to the CHARMM22 protein force field: dynamics of hen lysozyme, *Biophys. J.*, vol. 90, pp. L36–L37, 2006.

[105] Guvench, O., Qu, C. K., and Jr., A. D. M., Tyr66 acts as a conformational switch in the closed-to-open transition of the SHP-2 N-SH2-domain phophotyrosine peptide binding cleft, *BMC Struct. Biol.*, vol. 7, p. 14, 2007.

[106] Aleksandrov, A. and Simonson, T. A., A molecular mechanics model for imatinib and imatinib: kinase binding, *J. Comput. Chem.*, vol. 31, pp. 1550–1560, 2010.

[107] Gleeson, M. P., Deechongkit, S., and Ruchirawat, S., Molecular dynamics investgation of psalmopeotoxin I. Probing the relationship between 3D structure, antimalarial activity and thermal stability, *J. Molec. Mod.*, vol. 17, pp. 769–775, 2010.

[108] Liu, Y., Hsin, J., Kim, H., Selvin, P. R., and Schulten, K., Extension of a three-helix bundle domain of Myosin VI and key role of Calmodulins, *Biophys. J.*, vol. 100, pp. 2964–2973, 2011.

[109] Cino, E. A., Choy, W.-Y., and Karttunen, M., Comparison of Secondary Structure Formation Using 10 Different Force Fields in Microsecond Molecular Dynamics Simulations, *J. Chem. Theory Comput.*, vol. 8, pp. 2725–2740, 2012.

[110] Best, R. B. and Hummer, G., Optimized Moelcular Dynamics Force Fields Applied to the Helix-Coil Transition of Polypeptides, *J. Phys. Chem. B*, vol. 113, pp. 9004–9015, 2009.

[111] Mittal, J. and Best, R. B., Tackling Force-Field Bias in Protein Folding Simulations: Folding of Villin HP35 and Pin WW Domains in Explicit Water, *Biophys. J.*, vol. 99, pp. L26–L28, 2010.

[112] Best, R. B. and Mittal, J., Balance Between alpha and beta Structures in Ab Initio Protein, *J. Phys. Chem. B*, vol. 114, pp. 8790–8798, 2010.

[113] Best, R. B., Zhu, X., Shim, J., Lopes, P. E. M., Mittal, J., Feig, M., and MacKerell, Jr., A. D., Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone phi, psi and Side-Chain chi(1) and chi(2) Dihedral Angles, *J. Chem. Theory Comput.*, vol. 8, no. 9, pp. 3257–3273, 2012.

[114] Zhu, X., Lopes, P. E. M., Shim, J., and Alexander D. Mackerell, J., Correction to Intrinsic Energy Landscapes of Amino Acid Side-Chains, *J. Chem. Inf. Model.*, vol. 52, no. 8, pp. 2317–2318, 2012.

[115] Roterman, I. K., Gibson, K. D., and Scheraga, H. A., A COMPARISON OF THE CHARMM, AMBER AND ECEPP POTENTIALS FOR PEPTIDES .1. CONFORMATIONAL PREDICTIONS FOR THE TANDEMLY REPEATED PEPTIDE (ASN-ALA-ASN-PRO)9, *J. Biomol. Struct. Dyn.*, vol. 7, no. 3, pp. 391–419, 1989.

[116] Roterman, I. K., Lambert, M. H., Gibson, K. D., and Scheraga, H. A., A COMPARISON OF THE CHARMM, AMBER AND ECEPP POTENTIALS FOR PEPTIDES .2. PHI-PSI MAPS FOR N-ACETYL ALANINE N'-METHYL AMIDE - COMPARISONS, CONTRASTS AND SIMPLE EXPERIMENTAL TESTS, *J. Biomol. Struct. Dyn.*, vol. 7, no. 3, pp. 421–453, 1989.

[117] Hobza, P., Hubalek, F., Kabelac, M., Mejzlik, P., Sponer, J., and Vondrasek, J., Ability of empirical potentials (AMBER, CHARMM, CVFF, OPLS, Poltev) and semi-empirical quantum chemical methods (AM1, MNDO/M, PM3) to describe H-bonding in DNA base pairs; Comparison with ab initio results, *Chem. Phys. Lett.*, vol. 257, no. 1, pp. 31–35, 1996.

[118] Mu, Y. G., Kosov, D. S., and Stock, G., Conformational dynamics of trialanine in water. 2. Comparison of AMBER, CHARMM, GROMOS, and OPLS force fields to NMR and infrared experiments, *J. Phys. Chem. B*, vol. 107, no. 21, pp. 5064–5073, 2003.

[119] Martin, M. G., Comparison of the AMBER, CHARMM, COMPASS, GROMOS, OPLS, TraPPE and UFF force fields for prediction of vapor-liquid coexistence curves and liquid densities, *Fluid Phase Equilib.*, vol. 248, no. 1, pp. 50–55, 2006.

[120] Caleman, C., van Maaren, P. J., Hong, M., Hub, J. S., Costa, L. T., and van der Spoel, D., Force Field Benchmark of Organic Liquids: Density, Enthalpy of Vaporization, Heat Capacities, Surface Tension, Isothermal Compressibility, Volumetric Expansion Coefficient, and Dielectric Constant, *J. Chem. Theory Comput.*, vol. 8, no. 1, pp. 61–74, 2012.

[121] Debiec, K. T., Gronenborn, A. M., and Chong, L. T., Evaluating the Strength of Salt Bridges: A Comparison of Current Biomolecular Force Fields, *J. Phys. Chem. B*, vol. 118, no. 24, pp. 6561–6569, 2014.

[122] Smith, M. D., Rao, J. S., Segelken, E., and Cruz, L., Force-Field Induced Bias in the Structure of A beta(21-30): A Comparison of OPLS, AMBER, CHARMM, and GROMOS Force Fields, *J. Chem. Inf. Model.*, vol. 55, no. 12, pp. 2587–2595, 2015.

[123] Brown, L. Y. and Brown, S. A., Alanine tracts: the expanding story of human illness and trinucleotide repeats, *Trends Genet.*, vol. 20, no. 1, pp. 51–58, 2004.

[124] Brais, B., Bouchard, J. P., Xie, Y. G., Rochefort, D. L., Chretien, N., Tome, F. M. S., Lafreniere, R. G., Rommens, J. M., Uyama, E., Nohira, O., Blumen, S., Korcyn, A. D., Heutink, P., Mathieu, J., Duranceau, A., Codere, F., Fardeau, M., and Rouleau, G. A., Short GCG expansions in the PABP2 gene cause oculopharyngeal muscular dystrophy, *Nature Genetics*, vol. 18, no. 2, pp. 164–167, 1998.

[125] Messaed, C. and Rouleau, G. A., Molecular mechanisms underlying polyalanine diseases, *Neurobiol. Dis.*, vol. 34, no. 3, pp. 397–405, 2009.

[126] Lopez, M. M., Chin, D.-H., Baldwin, R. L., and Makhatadze, G. I., The enthalpy of the alanine peptide helix measured by isothermal titration calorimetry using metal-binding to induce helix formation, *Proc. Natl. Acad. Sci. U.S.A.*, vol. 99, pp. 1298–1302, 2002.

[127] Scholtz, J. M., Marqusee, S., Baldwin, R. L., York, E. J., Stewart, J. M., Santoro, M., and Bolen, D. W., Calorimetric determination of the enthalpy change for the α-helix to coil transition of an alanine peptide in water, *Proc. Natl. Acad. Sci. U.S.A.*, vol. 88, no. 7, pp. 2854–2858, 1991.

[128] Rialdi, G. and Hermans Jr., J., Calorimetric heat of the helix-coil transition of poly-L-glutamic acid, *J. Am. Chem. Soc.*, vol. 88, no. 24, pp. 5719–5720, 1966.

[129] Pace, C. N. and Scholtz, J. M., A helix propensity scale based on experimental studies of peptides and proteins, *Biophys. J.*, vol. 75, no. 1, pp. 422–427, 1998.

[130] Palencar, P. and Bleha, T., Folding of polyalanine into helical hairpins, *Macromol. Theory Simul.*, vol. 19, no. 8, pp. 488–495, 2010.

[131] Levy, Y., Jortner, J., and Becker, O. M., Solvent effects on the energy landscapes and folding kinetics of polyalanine, *Proc. Natl. Acad. Sci. U.S.A.*, vol. 98, no. 5, pp. 2188–2193, 2001.

[132] Mortenson, P. N., Evans, D. A., and Wales, D. J., Energy landscapes of model polyalanines, *J. Chem. Phys.*, vol. 117, pp. 1363–1376, 2002.

[133] Job, G. E., Kennedy, R. J., Heitmann, B., Miller, J. S., Walker, S. M., and Kemp, D. S., Temperature- and length-dependent energetics of formation for polyalanine helices in water: assignment of w(Ala)(n,T) and temperature-dependent CD ellipticity etandards, *J. Am. Chem. Soc.*, vol. 128, no. 25, pp. 8227–8233, 2006.

[134] NategholEslam, M., Holland, B. W., Gray, C. G., and Tomberli, B., Drift-oscillatory steering with the forward-reverse method for calculating the potential of mean force, *Phys. Rev. E*, vol. 83, p. 021114, 2011.

[135] Scheraga, H. A., Vila, J. A., and Ripoll, D. R., Helix-coil transitions re-visited, *Biophys. Chem.*, vol. 98, no. 101, pp. 255–265, 2002.

[136] Spek, E. J., Olson, C. A., Shi, Z., and Kallenbach, N. R., Alanine is an intrinsic alpha-helix stabilizing amino acid, *J. Am. Chem. Soc.*, vol. 121, no. 23, pp. 5571–5572, 1999.

[137] Yang, S. and Cho, M., Thermal denaturation of polyalanine peptide in water by molecular dynamics simulations and theoretical prediction of infrared spectra: Helix-coil transition kinetics, *J. Phys. Chem. B*, vol. 111, no. 3, pp. 605–617, 2007.

[138] Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L., Comparison of simple potential functions for simulating liquid water, *J. Chem. Phys.*, vol. 79, pp. 926–935, 1983.

[139] Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kale, L., and Schulten, K., Scalable Molecular Dynamics with NAMD, *J. Comput. Chem.*, vol. 26, pp. 1781–1802, 2005.

[140] Biron, Z., Khare, S., Samson, A. O., Hayek, Y., Naider, F., and Anglister, J., A monomeric 3(10)-helix is formed in water by a 13-residue peptide representing the neutralizing determinant of HIV-1 on gp41, *Biochemistry*, vol. 41, no. 42, pp. 12687–12696, 2002.

[141] Barbato, G., Bianchi, E., Ingallinella, P., Hurni, W. H., Miller, M. D., Cilibertol, G., Cortese, R., Bazzo, R., Shiver, J. W., and Pessi, A., Structural analysis of the epitope of the Anti-HIV antibody 2F5 sheds light into its mechanism of neutralization and HIV fusion, *J. Mol. Biol.*, vol. 330, no. 5, pp. 1101–1115, 2003.

[142] Bryson, S., Cunningham, A., Ho, J., Hynes, R. C., Isenman, D. E., Barber, B. H., Kunert, R., Katinger, H., Klein, M., and Pai, E. F., Cross-neutralizing human monoclonal anti-HIV-1 antibody 2F5: Preparation and crystallographic analysis of the free and epitope-complexed forms of its F-ab' fragment, *Protein Pept. Lett.*, vol. 8, no. 5, pp. 413–418, 2001.

[143] Gregor, C. R., Cerasoli, E., Tulip, P. R., Ryadnov, M. G., Martyna, G. J., and Crain, J., Autonomous folding in the membrane proximal HIV peptide gp41(659-671): pH tuneability at micelle interfaces, *Phys. Chem. Chem. Phys.*, vol. 13, no. 1, pp. 127–135, 2011.

[144] Kulik, H. J., Luehr, N., Ufimtsev, I. S., and Martinez, T. J., Ab Initio Quantum Chemistry for Protein Structures, *J. Phys. Chem. B*, vol. 116, no. 41, pp. 12501–12509, 2012.

[145] Zhang, Y. and Sagui, C., The gp41(659-671) HIV-1 Antibody Epitope: A Structurally Challenging Small Peptide, *J. Phys. Chem. B*, vol. 118, no. 1, pp. 69–80, 2014.

[146] Cerasoli, E., Ravi, J., Gregor, C., Hussain, R., Siligardi, G., Martyna, G., Crain, J., and Ryadnov, M. G., Membrane mediated regulation in free peptides of HIV-1 gp41: minimal modulation of the hemifusion phase, *Phys. Chem. Chem. Phys.*, vol. 14, no. 3, pp. 1277–1285, 2012.

[147] Scholtz, J. M., Qian, H., York, E. J., Stewart, J. M., and Baldwin, R. L., Parameters of Helix-Coil transition theory for alanine-based peptides of varying chain length in water, *Biopolymers*, vol. 31, no. 13, pp. 1463–1470, 1991.

[148] Sheinerman, F. B. and Brooks III, C. L., Helices in peptides and proteins as studied by modified Zimm-Bragg theory, *J. Am. Chem. Soc.*, vol. 117, pp. 10098–10103, 1995.

[149] Millhauser, G. L., Stenland, C. J., Hanson, P., Bolin, K. A., and van de Ven, F. J. M., Estimating the realtive populations of 3_{10}-helix and α-helix in Ala-rich peptides: A hydrogen exchange and high field NMR study, *J. Mol. Biol.*, vol. 267, pp. 963–974, 1997.

[150] Barlow, D. J. and Thornton, J. M., Helix geometry in protiens, *J. Mol. Biol.*, vol. 201, pp. 601–619, 1988.

[151] Young, W. S. and Brooks III, C. L., A microscopic view of helix propagation: N and C-terminal helix growth in alanine helices, *J. Mol. Biol.*, vol. 259, pp. 560–572, 1996.

[152] Bryant, Z., Pande, V. S., and Rokhsar, D. S., Mechanical Unfolding of a Beta-Hairpin Using Molecular Dynamics, *Biophys. J.*, vol. 78, no. 2, pp. 584–589, 2000.

[153] Soman, K. V., Karimi, A., and Case, D. A., Unfolding of an α-helix in water, *Biopolymers*, vol. 31, no. 12, pp. 1351–1361, 1991.

[154] Tirado-Rives, J. and Jorgensen, W. L., Molecular Dynamic Simulations of the Unfolding of an α-helical Analogue of Ribonuclease A S-Peptide in Water, *Biochemistry*, vol. 30, no. 16, pp. 3864–3871, 1991.

[155] Marqusee, S. and Baldwin, R., Helix stabilization by $Glu^{-} \cdots Lys^{+}$ salt bridges in short peptides of de novo design, *Proc. Natl. Acad. Sci. U.S.A.*, vol. 84, pp. 8898–8902, 1987.

[156] Bustamante, C., Chemla, Y. R., Forde, N. R., and Izhaky, D., Mechanical Processes in Biochemistry, *Annu. Rev. Biochem.*, vol. 73, no. 1, pp. 705–748, 2004.

[157] Karplus, M. and McCammon, J. A., Molecular Dynamics Simulations of Biomolecules, *Nat. Struct. Biol.*, vol. 9, no. 9, pp. 646–652, 2002.

[158] Zimm, B. H. and Bragg, J. K., Theory of the Phase Transition Between Helix and Random Coil in Polypeptide Chains, *J. Chem. Phys.*, vol. 31, no. 2, pp. 526–535, 1959.

[159] Muñoz, V., Thompson, P. A., Hofrichter, J., and Eaton, W. A., Folding Dynamics and Mechanism of Beta-Hairpin Formation, *Nature (London)*, vol. 390, pp. 196–199, 1997.

[160] Hughes, R. M. and Waters, M. L., Model Systems for Beta-Hairpins and Beta-Sheets, *Curr. Opin. Struct. Biol.*, vol. 16, no. 4, pp. 514–524, 2006.

[161] Kolinski, A., Ilkowski, B., and Skolnick, J., Dynamics and Thermodynamics of Beta-Hairpin Assembly: Insights from Various Simulation Techniques, *Biophys. J.*, vol. 77, no. 6, pp. 2942–2952, 1999.

[162] Dinner, A. R., Lazaridis, T., and Karplus, M., Understanding Beta-Hairpin Formation, *Proc. Natl. Acad. Sci. U.S.A.*, vol. 96, pp. 9068–9073, 1999.

[163] Du, D., Zhu, Y., Huang, C.-Y., and Gai, F., Understanding the Key Factors that Control the Rate of Beta-Hairpin Folding, *Proc. Natl. Acad. Sci. U.S.A.*, vol. 101, no. 45, pp. 15915–15920, 2004.

[164] Best, R. M. and Mittal, J., Microscopic Events in Beta-Hairpin Folding from Alternative Unfolded Ensembles, *Proc. Natl. Acad. Sci. U.S.A.*, vol. 108, no. 27, pp. 11087–11092, 2011.

[165] Kim, J. and Keiderling, T. A., All-Atom Molecular Dynamics Simulations of Beta-Hairpins Stabilized by a Tight Turn: Pronounced Heterogeneous Folding Pathways, *J. Phys. Chem. B*, vol. 114, pp. 8494–8504, 2010.

[166] Cochran, A. G., Skelton, N. J., and Starovasnik, M. A., Tryptophan Zippers: Stable, Monomeric Beta-Hairpins, *Proc. Natl. Acad. Sci. U.S.A.*, vol. 98, no. 10, pp. 5578–5583, 2001.

[167] Honda, S., Yamasaki, K., Sawada, Y., and Morii, H., 10 Residue Folded Peptide Designed by Segment Statistics, *Structure*, vol. 12, no. 8, pp. 1507–1518, 2004.

[168] Honda, S., Akiba, T., Kato, Y. S., Sawada, Y., Sekijima, M., Ishimura, M., Ooishi, A., Watanabe, H., Odahara, T., and Harata, K., Crystal Structure of a Ten-Amino Acid Protein, *J. Am. Chem. Soc.*, vol. 130, no. 46, pp. 15237–15331, 2008.

[169] Neidigh, J. W., Fesinmeyer, R. M., and Andersen, N. H., Designing a 20-Residue Protein, *Nat. Struct. Biol.*, vol. 9, pp. 425–430, 2002.

[170] Hopping, G., Kellock, J., Caughey, B., and Daggett, V., Designed Trpzip-3 Beta-Hairpin Inhibits Amyloid Formation in Two Different Amyloid Systems, *ACS Med. Chem. Lett.*, vol. 4, no. 9, pp. 824–828, 2013.

[171] Yu, Z., Selvam, S., and Mao, H., Intermediates Stabilized by Tryptophan Pairs Exist in Beta-Hairpins, *Biochemistry*, vol. 53, no. 38, pp. 5978–5986, 2014.

[172] Roitberg, A. E., Okur, A., and Simmerling, C., Coupling of Replica Exchange Simulations to a Non-Boltzmann Structure Reservoir, *J. Phys. Chem. B*, vol. 111, no. 10, pp. 2415–2418, 2007.

[173] Doshi, U. and Hamelberg, D., Achieving Rigorous Accelerated Conformational Sampling in Explicit Solvent, *J. Phys. Chem. Lett.*, vol. 5, no. 7, pp. 1217–1224, 2014.

[174] Miao, Y., Sinko, W., Pierce, L., Bucher, D., Walker, R. C., and McCammon, J. A., Improved Reweighting of Accelerated Molecular Dynamics Simulations for Free Energy Calculation, *J. Chem. Theory Comput.*, vol. 10, no. 7, pp. 2677–2689, 2014.

[175] Ramirez, C. L., Zeida, A., Jara, G. E., Roitberg, A. E., and Marti, M. A., Improving Efficiency in SMD Simulations Through a Hybrid Differential Relaxation Algorithm, *J. Chem. Theory Comput.*, vol. 10, no. 10, pp. 4609–4617, 2014.

[176] Coifman, R. R., Lafon, S., Lee, A. B., Maggioni, M., Nadler, B., Warner, F., and Zucker, S. W., Geometric Diffusions as a Tool for Harmonic Analysis and Structure Definition of Data: Diffusion Maps, *Proc. Natl. Acad. Sci. U.S.A.*, vol. 102, no. 21, pp. 7426–7431, 2005.

[177] Kim, S. B., Dsilva, C. J., Kevrekidis, I. G., and Debenedetti, P. G., Systematic Characterization of Protein Folding Pathways Using Diffusion Maps: Application to Trp-Cage Miniprotein, *J. Chem. Phys.*, vol. 142, no. 8, p. 085101, 2015.

[178] Settanni, G., Serquera, D., Marszalek, P. E., Paci, E., and Itzhaki, L. S., Effects of Ligand Binding on the Mechanical Properties of Ankyrin Repeat Protein Gankyrin, *PLoS Comp. Biol.*, vol. 9, no. 1, p. e1002864, 2013.

[179] Zerze, G. H., Uz, B., and Mittal, J., Folding Thermodynamics of Beta-Hairpins Studied by Replica-Exchange Molecular Dynamics Simulations, *Proteins: Struct., Func., Bioinf.*, vol. 83, no. 7, pp. 1307–1315, 2015.

[180] Snow, C. D., Qiu, L., Du, D., Gai, F., Hagen, S. J., and Pande, V. S., Trp Zipper Folding Kinetics by Molecular Dynamics and Temperature-Jump Spectroscopy, *Proc. Natl. Acad. Sci. U.S.A.*, vol. 101, pp. 4077–4082, 2004.

[181] Settanni, G. and Fersht, A. R., High Temperature Unfolding Simulations of the TRPZ1 Peptide, *Biophys. J.*, vol. 94, pp. 4444–4453, 2008.

[182] Kusumaatmaja, H., Whittleston, C. S., and Wales, D. J., A Local Rigid Body Framework for Global Optimization of Biomolecules, *J. Chem. Theory Comput.*, vol. 8, pp. 5159–5165, 2012.

[183] Gellman, S. H., Minimal model systems for beta sheet secondary structure in proteins, *Curr. Opin. Chem. Biol.*, vol. 2, no. 6, pp. 717–725, 1998.

[184] Takekiyo, T., Wu, L., Yoshimura, Y., Shimizu, A., and Keiderling, T. A., Relationship between hydrophobic interactions and secondary structure stability for trpzip beta-hairpin peptides, *Biochemistry*, vol. 48, no. 7, pp. 1543–1552, 2009.

[185] Ciani, B., Jourdan, M., and Searle, M. S., Stabilization of Beta-Hairpin Peptides by Salt Bridges: Role of Preorganization in the Energetic Contribution of Weak Interactions, *J. Am. Chem. Soc.*, vol. 125, no. 30, pp. 9038–9047, 2003.

[186] Missimer, J. H., Steinmetz, M. O., Baron, R., Winkler, F. K., Kammerer, R. A., Daura, X., and van Gunsteren, W. F., Configurational entropy elucidates the role of salt-bridge networks in protein thermostability, *Protein Sci.*, vol. 16, no. 7, pp. 1349–1359, 2007.

[187] Kuo, H.-T., Fang, C.-J., Tsai, H.-Y., Yang, M.-F., Chang, H.-C., Liu, S.-L., Kuo, L.-H., Wang, W.-R., Yang, P.-A., Huang, S.-J., Huang, S.-L., and Cheng, R. P., Effect of Charged Amino Acid Side Chain Length on Lateral Cross-Strand Interactions between Carboxylate-Containing Residues and Lysine Analogues in a Beta-Hairpin, *Biochemistry*, vol. 52, no. 51, pp. 9212–9222, 2013.

[188] Kiehna, S. E. and Waters, M. L., Sequence dependence of beta-hairpin structure: Comparison of a salt bridge and an aromatic interaction, *Protein Sci.*, vol. 12, no. 12, pp. 2657–2667, 2003.

[189] Sheridan, R. P., Lee, R. H., Peters, N., and Allen, L. C., Hydrogen-bond cooperativity in protein secondary structure, *Biopolymers*, vol. 18, no. 10, pp. 2451–2458, 1979.

[190] Nowick, J. S., Exploring beta-Sheet Structure and Interactions with Chemical Model Systems, *Acc. Chem. Res.*, vol. 41, no. 10, pp. 1319–1330, 2008.

[191] Wu, Y.-D. and Zhao, Y.-L., A theoretical study on the origin of cooperativity in the formation of 3(10)- and alpha-helices, *J. Am. Chem. Soc.*, vol. 123, no. 22, pp. 5313–5319, 2001.

[192] Zhao, Y.-L. and Wu, Y.-D., A theoretical study of beta-sheet models: Is the formation of hydrogen-bond networks cooperative?, *J. Am. Chem. Soc.*, vol. 124, no. 8, pp. 1570–1571, 2002.

[193] Waldburger, C. D., Schildbach, J. F., and Sauer, R. T., Are buried salt bridges important for protein stability and conformational specificity?, *Nat. Struct. Biol.*, vol. 2, no. 2, pp. 122–128, 1995.

[194] Maynard, A. J., Sharman, G. J., and Searle, M. S., Origin of beta-hairpin stability in solution: Structural and thermodynamic analysis of the folding of model peptide supports hydrophobic stabilization in water, *J. Am. Chem. Soc.*, vol. 120, no. 9, pp. 1996–2007, 1998.

[195] Robinson, A. C., Castaneda, C. A., Schlessman, J. L., and Garcia-Moreno, B. E., Structural and thermodynamic consequences of burial of an artificial ion pair in the hydrophobic interior of a protein, *Proc. Natl. Acad. Sci. U.S.A.*, vol. 111, no. 32, pp. 11685–11690, 2014.

[196] Olsen, K. A., Fesinmeyer, R. M., Stewart, J. M., and Andersen, N. H., Hairpin Folding Rates Reflect Mutations Within and Remote from the Turn Region, *Proc. Natl. Acad. Sci. U.S.A.*, vol. 102, no. 43, pp. 15483–15487, 2005.

[197] Herzberg, O. and Moult, J., Analysis of the steric strain in the polypeptide backbone of protein molecules, *Proteins: Struct. Func. Gen.*, vol. 11, no. 3, pp. 223–229, 1991.

[198] de Graff, A. M. R., Shannon, G., Farrell, D. W., Williams, P. M., and Thorpe, M. F., Protein unfolding under force: Crack propagation in a network, *Biophys. J.*, vol. 101, no. 3, pp. 736–744, 2011.

[199] Ma, B. and Nussinov, R., Molecular dynamics simulations of a beta-hairpin fragment of protein G: Balance between side-chain and backbone forces, *J. Mol. Biol.*, vol. 296, no. 4, pp. 1091–1104, 2000.

[200] Chiti, F., Stefani, M., Taddei, N., Ramponi, G., and Dodson, C. M., Rationalization of the effects of mutations on peptide and protein aggregation rates, *Nature (London)*, vol. 424, no. 6950, pp. 805–808, 2003.

[201] Chen, W., Shi, C., MacKerell, Jr., A. D., and Shen, J., Conformational Dynamics of Two Natively Unfolded Fragment Peptides: Comparison of the AMBER and CHARMM Force Fields, *J. Phys. Chem. B*, vol. 119, no. 25, pp. 7902–7910, 2015.

[202] Lee, J. and Shin, S., Understanding Beta-Hairpin Formation by Molecular Dynamics Simulations of unfolding, *Biophys. J.*, vol. 81, no. 5, pp. 2507–2516, 2001.

[203] Bolhuis, P. G., Kinetic Pathways of Beta-Hairpin (Un)folding in Explicit Solvent, *Biophys. J.*, vol. 88, no. 1, pp. 50–61, 2005.

[204] Freddolino, P. L., Lui, F., Gruebele, M., and Schulten, K., Ten-Microsecond Molecular Dynamics Simulation of a Fast-Folding WW Domain, *Biophys. J.*, vol. 94, no. 10, pp. L75–L77, 2008.

[205] Best, R. B., Mittal, J., Feig, M., and MacKerell, Jr., A. D., Inclusion of Many-Body Effects in the Additive CHARMM Protein CMAP Potential Results in Enhanced Cooperativity of Alpha-Helix and Beta-Hairpin Formation, *Biophys. J.*, vol. 103, no. 5, pp. 1045–1051, 2012.

[206] Minor Jr., D. L. and Kim, P. S., Measurement of the beta-sheet-forming propensities of amino acids, *Nature (London)*, vol. 367, no. 6464, pp. 660–663, 1994.

[207] Smith, C. K. and Regan, L., Guidelines for protein design - The energetics of beta-sheet side-chain interactions, *Science*, vol. 270, no. 5238, pp. 980–982, 1995.

[208] Phillips, S. T., Piersanti, G., and Bartlett, P. A., Quantifying amino acid conformational preferences and side-chain-side-chain interactions in beta-hairpins, *Proc. Natl. Acad. Sci. U.S.A.*, vol. 102, no. 39, pp. 13737–13742, 2005.

[209] Becktel, W. J. and Schellman, J. A., Protein stability curves, *Biopolymers*, vol. 26, no. 11, pp. 1859–1877, 1987.

[210] Bolhuis, P. G., Transition-Path Sampling of Beta-Hairpin Folding, *Proc. Natl. Acad. Sci. U.S.A.*, vol. 100, no. 21, pp. 12129–12134, 2003.

[211] Nymeyer, H., Energy Landscape of the Trpzip2 Peptide, *J. Phys. Chem. B*, vol. 113, no. 24, pp. 8288–8295, 2009.

[212] Zhang, J., Qin, M., and Wang, W., Folding Mechanism of Beta-Hairpins Studied by Replica Exchange Molecular Simulations, *Proteins: Struct., Func., Bioinf.*, vol. 62, no. 3, pp. 672–685, 2006.

[213] Zhao, G.-J. and Cheng, C.-L., Molecular Dynamics Simulation Exploration of Unfolding and Refolding of a Ten-Amino Acid Miniprotein, *Amino Acids*, vol. 43, pp. 557–565, 2012.

[214] Smith, A. W., Lessing, J., Ganim, Z., Peng, C. S., Tokmakoff, A., Roy, S., Jansen, T. L. C., and Knoester, J., Melting of a Beta-Hairpin Peptide Using Isotope-Edited 2D IR Spectroscopy and Simulations, *J. Phys. Chem. B*, vol. 114, no. 34, pp. 10913–10924, 2010.

[215] Radford, I. H., Fersht, A. R., and Settanni, G., Combination of Markov State Models and Kinetic Networks for the Analysis of Molecular Dynamics Simulations of Peptide Folding, *J. Phys. Chem. B*, vol. 115, no. 22, pp. 7459–7471, 2011.

[216] Terada, T., Satoh, D., Mikawa, T., Ito, Y., and Shimizu, K., Understanding the Roles of Amino Acid Residues in Tertiary Structure Formation of Chignolin by Using Molecular Dynamics, *Proteins: Struct., Func., Bioinf.*, vol. 73, no. 3, pp. 621–631, 2008.

[217] Hatfield, M. P. D., Murphy, R. F., and Lovas, S., Molecular Dynamics Analysis of the Conformations of a Beta-Hairpin Miniprotein, *J. Phys. Chem. B*, vol. 114, no. 8, pp. 3028–3037, 2010.

[218] Distefano, M. D., Zhong, A., and Cochran, A. G., Quantifying beta-sheet stability by phage display, *J. Mol. Biol.*, vol. 322, no. 1, pp. 179–188, 2002.

[219] Wu, L., McElheny, D., Huang, R., and Keiderling, T. A., Role of tryptophan-tryptophan interactions in trpzip beta-hairpin formation, structure, and stability, *Biochemistry*, vol. 48, pp. 10362–10371, 2009.

[220] Kelly, S. M. and Price, N. C., The application of circular dichroism to studies of protein folding and unfolding, *Biochim. Biophys. Acta, Protein Struct. Mol. Enzymol.*, vol. 1338, no. 2, pp. 161–185, 1997.

[221] Kelly, S. M., Jess, T. J., and Price, N. C., How to study proteins by circular dichroism, *Biochim. Biophys. Acta, Proteins Proteomics*, vol. 1751, no. 2, pp. 119–139, 2005.

[222] Ranjbar, B. and Gill, P., Circular Dichroism Techniques: Biomolecular and Nanostructural Analyses- A Review, *Chem. Biol. Drug Des.*, vol. 74, no. 2, pp. 101–120, 2009.

[223] Mukherjee, S., Chowdhury, P., and Gai, F., Tuning the Cooperativity of the Helix-Coil Transition by Aqueous Reverse Micelles, *J. Phys. Chem. B*, vol. 110, no. 24, pp. 11615–11619, 2006.

[224] Schulz, J. C. F., Miettinen, M. S., and Netz, R. R., Unfolding and Folding Internal Friction of Beta-Hairpins is Smaller than that of Alpha-Helices, *J. Phys. Chem. B*, vol. 119, no. 13, pp. 4565–4574, 2015.

[225] Stranger, H. E., Syud, F. A., Espinosa, J. F., Giriat, I., Muir, T., and Gellman, S. H., Length-Dependent Stability and Strand Length Limits in Antiparallel Beta-Sheet Secondary Structure, *Proc. Natl. Acad. Sci. U.S.A.*, vol. 98, no. 21, pp. 12015–12020, 2001.

[226] Monks, S. A., Karagianis, G., Howlett, G. J., and Norton, R. S., Solution structure of human neuropeptide Y, *J.Biomol.NMR*, vol. 8, pp. 379–390, 1996.

[227] Larhammar, D., Structural diversity of receptors for neuropeptide Y, peptide YY and pancreatic polypeptide, *Regul. Pept.*, vol. 65, pp. 165–174, 1996.

[228] Turton, M. D., O'Shea, D., and Bloom, S. R., *Central effects of neuropeptide Y with emphasis on its role in obesity and diabetes*, pp. 15–39. San Diego, CA: Academic Press, 1997.

[229] Dyzma, M., Boudjeltia, K. Z., Faraut, B., and Kerkhofs, M., Neuropeptide Y and sleep, *Sleep Med. Rev.*, vol. 14, no. 3, pp. 161–165, 2010.

[230] Robertson, C. R., Flynn, S. P., White, H. S., and Bulaj, G., Anticonvulsant neuropeptides as drug leads for neurological diseases, *Nat. Prod. Rep.*, vol. 28, no. 4, pp. 741–762, 2011.

[231] Walther, C., Moerl, K., and Beck-Sickinger, A. G., Neuropeptide Y receptors: ligand binding and trafficking suggest novel approaches in drug development, *J. Pep. Sci.*, vol. 17, no. 4, pp. 233–246, 2011.

[232] Nordmann, A., Blommers, M. J. J., Fretz, H., Arvinte, T., and Drake, F., Aspects of the molecular structure and dynamics of neuropeptide Y, *Eur. J. Biochem.*, vol. 261, pp. 216–226, 1999.

[233] Bettio, A., Dinger, M. C., and Beck-Sickinger, A. G., The neuropeptide Y monomer in solution is not folded in the pancreatic-polypeptide fold, *Protein Sci.*, vol. 11, pp. 1834–1844, 2002.

[234] Neumoin, A., Mares, J., Lerch-Bader, M., Bader, R., and Zerbe, O., Probing the formation of stable tertiary structure in a model miniprotein at atomic resolution: Determinants of stability of a helical hairpin, *J. Am. Chem. Soc.*, vol. 129, no. 28, pp. 8811–8817, 2007.

[235] Gnandt, D., Utz, N., Blumen, A., and Koslowski, T., Protein displacements under external forces: An atomistic Langevin dynamics approach, *J. Chem. Phys.*, vol. 130, no. 8, p. 085104, 2009.

[236] Glover, M. S., Bellinger, E. P., Radivojac, P., and Clemmer, D. E., Penultimate Proline in Neuropeptides, *Anal. Chem.*, vol. 87, no. 16, pp. 8466–8472, 2015.

[237] Lohans, C. T., Towle, K. M., Miskolzie, M., McKay, R. T., van Belkum, M. J., McMullen, L. M., and Vederas, J. C., Solution structures of the linear leaderless bacteriocins Enterocin 7A and 7B resemble Carnocyclin A, a circlar antimicrobial peptide, *Biochemistry*, vol. 52, pp. 3987–3994, 2013.

[238] Riley, M. A. and Wertz, J. E., Bacteriocins: Evolution, ecology, and application, *Annu. Rev. Microbiol.*, vol. 56, no. 1, pp. 117–137, 2002.

[239] Greenfield, N. J., Using circular dichroism spectra to estimate protein secondary structure, *Nat Protoc.*, vol. 1, no. 6, pp. 28876–2890, 2006.

[240] Besenicar, M., Macek, P., Lakey, J. H., and Anderluh, G., Surface plasmon resonance in protein-membrane interactions, *Chem. Phys. Lipids*, vol. 141, no. 1, pp. 169–178, 2006.

[241] Statnick, M. A., Schober, D. A., Mayne, N. G., Burnett, J. P., and Gehlert, D. R., Analysis of NPY receptor subtypes in the human frontal cortex reveals abundant Y-1 mRNA and binding sites, *Peptides*, vol. 18, no. 1, pp. 137–143, 1997.

[242] Cowley, D. J., Hoflack, J. M., Pelton, J. T., and Saudek, V., Structure of neuropeptide Y dimer in solution, *Eur. J. Biochem.*, vol. 205, pp. 1099–1106, 1992.

[243] Mierke, D. F., Durr, H., Kessler, H., and Jung, G., Neuropeptide Y: Optimized solid-phase synthesis and conformational analysis in trifluoroethanol, *Eur. J. Biochem.*, vol. 206, pp. 39–48, 1992.

[244] Cole, J. L., Lary, J. W., Moody, T. P., and Laue, T. M., Analytical ultracentrifugation: Sedimentation velocity and sedimentation equilibrium, in *Biophysical tools for biologists: Vol. 1 in vitro techniques* (Correia, J. J. and Detrich, H. W., eds.), vol. 84, pp. 143–179, San Diego, CA: Elsevier Academic Press Inc., 2008.